Scientific Reports

https://doi.org/10.1038/s41598-026-53543-y

Article in Press

Hybrid multi-scale CNN-Residual-LSTM approach for robust state-of-charge estimation in lithium-ion batteries

Received: 1 December 2025

Accepted: 12 May 2026

Published online: 22 May 2026

Cite this article as: Oyucu S., Polat O., Türkoğlu M. *et al.* Hybrid multi-scale CNN-Residual-LSTM approach for robust state-of-charge estimation in lithium-ion batteries. *Sci Rep* (2026). https://doi.org/10.1038/s41598-026-53543-y

Saadin Oyucu, Onur Polat, Muammer Türkoğlu, Faruk Kürker, Ahmet Aksöz & Emre Biçer

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

ARTICLE IN PRESS

© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

ARTICLE IN PRESS

Hybrid multi-scale CNN-Residual-LSTM approach for robust state-of-charge estimation in lithium-ion batteries

Saadin Oyucu[1,*], Onur Polat[2], Muammer Türkoğlu[3], Faruk Kürker[4], Ahmet Aksöz[5],
Emre Biçer[6,*]

[1] Department of Computer Engineering, Faculty of Technology, Gazi University, 06560, Ankara, Türkiye; saadinoyucu@gazi.edu.tr

[2] Department of Computer Engineering, Bingöl University, Bingöl 12000, Turkey; opolat@bingol.edu.tr

[3] Department of Software Engineering, Faculty of Engineering and Natural Sciences, Samsun University, Samsun 55000, Türkiye; muammer.turkoglu@samsun.edu.tr

[4] Department of Electrical and Electronic Engineering, Faculty of Engineering, Adıyaman University, Adıyaman 02040, Türkiye; fkurker@adiyaman.edu.tr

[5] Department of Electrical-Electronics Engineering, Faculty of Engineering, Kayseri University, 38280, Kayseri, Türkiye; ahmetaksoz@kayseri.edu.tr

[6] Sivas Energy Storage Research and Application Center, Sivas University of Science and Technology, 58010, Sivas, Türkiye; emre.bicer@sivas.edu.tr

Corresponding Authors:

- Assoc. Prof. Dr. Saadin Oyucu
  saadinoyucu@gazi.edu.tr
- Prof. Dr. Emre Biçer
  emre.bicer@sivas.edu.tr

ARTICLE IN PRESS

Abstract

Accurate estimation of State of Charge (SoC) and State of Health (SoH) values in battery management systems is crucial for maintaining energy efficiency, ensuring operational safety, and sustaining battery life. This study proposes a hybrid deep learning (DL) architecture that combines multi-scale convolutional neural networks (CNNs), residual blocks, and long-short-term memory (LSTM) units. The proposed model is designed to learn both local spatial patterns and time-dependent relationships found in battery data. The developed model was trained using real battery datasets, and its performance was evaluated comparatively against traditional machine learning (ML) algorithms such as Random Forest (RF), Support Vector Regression (SVR), and Gradient Boosting (GBM). Model performance was analyzed using common regression metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and Coefficient of Determination (R^2). The results show

ARTICLE IN PRESS

that the proposed model achieves lower prediction errors compared to the methods evaluated (MAE = 0.1507, MSE = 0.2030, R^2 = 0.9998). Furthermore, scatter plots and residual distribution analyses reveal that the model predictions are quite close to the actual values and that the error distribution does not contain any significant systematic bias. Overall, the results demonstrate that the proposed hybrid architecture can produce accurate and stable results for SoC estimation and offers an approach that could contribute to the development of reliable battery management systems.

Keywords: Li-ion, State of Charge, CNN-Residual-LSTM, BMS

1 INTRODUCTION

Li-ion batteries (LIBs) are at the heart of the modern electrification system, being leveraged for electric vehicles (EVs), smart grids, consumer electronics, and renewable energy storage platforms. These have been widely accepted due to their higher energy densities, longer lifespans, and better charge-discharge efficiencies. On the other hand, under real working conditions, where extreme temperatures, rapid load changes, and fast charging come into play, all these put together necessitate being able to robustly estimate the internal state for safety, performance, and predictive maintenance [1-5].

Traditional model-based techniques such as equivalent circuit models (ECMs), Coulomb counting, and open-circuit voltage (OCV) methods often exhibit limited accuracy under complex operating conditions, mainly for the reason of their static parameters and propagation-based error generation [4, 6]. In recent years, data-driven DL frameworks have become increasingly prominent. Using large-scale operational datasets, these models readily discern the spatiotemporal interactions between electric and thermal variables. CNN and LSTM-based frameworks, unsurprisingly, carry out the process of extracting hierarchical features while modeling nonlinear degradation dynamics of the battery [1, 2, 7-13]

Development in the recent past led to advanced hybrid models, simultaneously comprising CNNs, Bidirectional Long Short-Term Memory Networks (BiLSTMs), attention mechanisms, residual connections, and metaheuristic optimizers such as genetic algorithms (GAs), particle swarm optimization (PSO), and wasserstein generative adversarial network (WGAN) augmentation [14-20]. These integrated frameworks provide improved adaptability to battery chemistries, C-rates, and ambient conditions. Noise reduction and emphasizing signal relevance were also ensured by preprocessing techniques such as complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN), grey relational analysis, and autoregressive integrated moving average (ARIMA) filtering to bring forth model robustness [18, 21-23].

Emerging trends further point to the use of transfer learning [24], uncertainty-aware estimation [2], graph-based architectures such as graph convolutional networks (GCNs) [25], along with joint multi-output learning

ARTICLE IN PRESS

[2, 26]. The increased profusion of public (e.g., NASA's prognostics center of excellence (NASA PCoE), center for advanced life cycle engineering (CALCE)) as well as proprietary electric vehicle (EV) datasets [6, 26, 27] allows further reproducibility and generalization on several use cases. Apart from purely data-driven approaches, recent studies have shown a growing interest in hybrid modeling approaches involving both electrochemical models and artificial intelligence. This class of AI-assisted electrochemical models combines the advantages of physically based models such as interpretation and consistency with those of data-driven models, namely adaptability and learning capabilities. Such a paradigm is able to identify dynamic parameters, ensure better tracking accuracy, and increase robustness under different conditions. Recent reviews demonstrate that the combination of physical models and machine learning can help bridge the gap between conventional modeling and black-box deep learning techniques in battery state estimation [28].

Filling this context gap, the present review study synthesizes 31 DL models proposed between 2020 and 2025 for the estimation of SoC, SoH, and remaining useful life (RUL). Table 1 is a comparative view based on the type of model, dataset properties, optimization procedure, and major findings. Such a systematic analysis can distill important trends in methodology and inform further research toward establishing robust and scalable battery diagnostic techniques.

Despite recent advances, three critical gaps currently remain in the DL-based battery estimation models: (1) fragmented architectural designs that confine themselves to either CNNs or LSTMs without proper integration of multi-scale spatial representation and residual refinement; (2) degraded generalization due to training against narrowly scoped datasets, commonly fixed to certain C-rates, chemistries, or temperature ranges; and (3) an established structured benchmarking in terms of model architectures, datasets, and performance metrics does not exist, which undermines reproducibility of results and their deployment into BMSs.

In response to such problems, a novel Multi-Scale CNN–Residual–LSTM architecture is proposed in this study to enable modeling of nonlinear battery degradation by integrating multi-resolution convolutional layers, Swish-activated residual blocks, and LSTM-based temporal modeling. The adopted validation strategy is based on the CALCE dataset, which was collected under controlled laboratory conditions using lithium cobalt oxide (LCO) batteries. While this dataset provides reliable and high-quality measurements, it does not include highly dynamic real-world driving cycles such as the urban dynamometer driving schedule (UDDS), US06 supplemental federal test procedure (US06), or beijing bus dynamic stress test (BBDST). Therefore, the generalizability of the model under real-world driving conditions remains a subject for future work. This limitation is clearly stated, and future work is planned to focus on validating the model using datasets that incorporate real-world dynamic driving conditions and different battery chemistries. Furthermore, the paper synthesizes a benchmark of 31 DL models from 2020

ARTICLE IN PRESS

to 2025, allowing for standardized comparison and guiding practical deployment. All these contributions lay down the foundations for interpretable, scalable, and real-time SoC prediction systems for next-generation energy storage. The framework thus stands for both methodological and practical advancement, where it is integrated into BMS environments.

Battery state estimation under different operating conditions presents significant challenges in terms of time series analysis. To address these challenges, Figure 1 presents a hybrid CNN–Residual–LSTM architecture that combines multi-scale convolutional feature extraction, residual-based representation learning using the Swish activation function, and long-term temporal modeling via LSTM layers. This integrated approach aims to overcome fundamental limitations observed in traditional CNN-based methods, such as the fragmented handling of spatial and temporal features and limited generalization ability under different usage profiles.

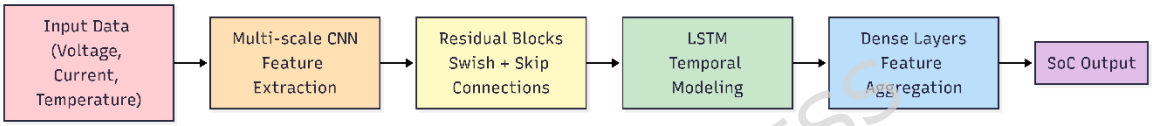


Figure 1. Overview of the proposed CNN–Residual–LSTM architecture for battery state of charge (SoC) estimation. The model combines multi-scale convolutional feature extraction, residual learning, and LSTM-based temporal modeling to produce the final SoC prediction.

Unlike many prior models that treat spatial and temporal aspects separately, this architecture offers an integrated learning mechanism that enhances both local sensitivity and long-range sequence interpretation. Its design supports stable performance even under fluctuating thermal or operational regimes, making it potentially applicable to real-world battery management systems, subject to further validation under realistic operating conditions. By balancing short-term feature responsiveness with long-term degradation tracking, the model contributes to more resilient and accurate SoC predictions.

Given the wide range of DL architectures proposed for battery diagnostics, there is a growing need for a structured taxonomy that clarifies design trends and methodological gaps. Figure 2 addresses this need by organizing 31 recent models into four major groups: CNN–LSTM hybrids, architectures incorporating attention or residual elements, graph-based frameworks, and models emphasizing preprocessing techniques such as CEEMDAN or ARIMA. This classification not only reveals the dominant modeling directions but also provides context for the architectural choices made in the present study.

ARTICLE IN PRESS

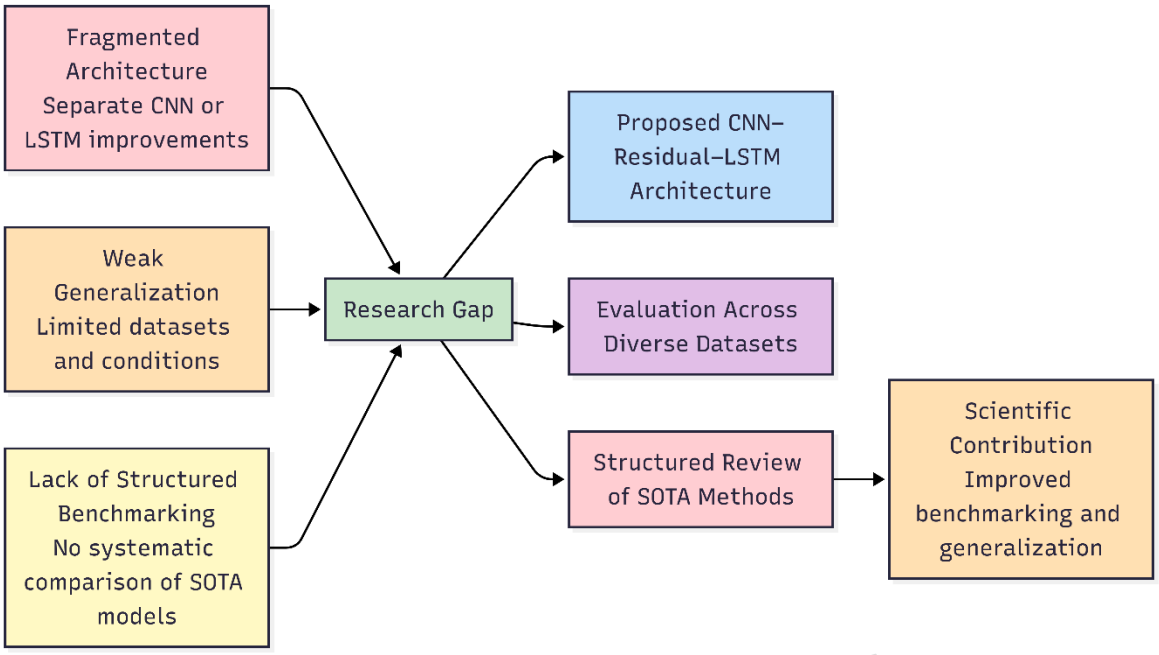


Figure 2. A conceptual framework demonstrating the research gaps in existing deep learning approaches for battery diagnosis and how the proposed CNN–Residual–LSTM architecture addresses these limitations through integrated spatiotemporal modeling and multi-dataset evaluation.

The visual comparison reinforces the rationale behind the model design adopted in this research. By bridging the spatial–temporal divide, improving adaptability to different battery datasets, and enabling systematic benchmarking, the proposed model directly responds to documented weaknesses in the field. This strategic positioning supports its relevance not only as a new architectural solution but also as a framework capable of unifying disparate modeling approaches under a more generalizable and testable structure.

The main contributions of this study can be summarized as follows:

- A novel hybrid deep learning architecture that integrates multi-scale convolutional feature extraction, residual learning with Swish activation, and LSTM-based temporal modeling within a unified framework for battery state estimation is proposed.
- Unlike conventional CNN–LSTM or ResNet–LSTM approaches, the proposed model uses a multi-scale convolutional structure with kernel sizes of 1, 3, and 5 combined with an enhanced residual block design. This enables the model to more effectively capture both local and global feature patterns.
- A systematic and controlled ablation study is conducted under identical experimental conditions to evaluate the individual contributions of the CNN, residual, and LSTM components, providing a clear and interpretable assessment of the model architecture.

ARTICLE IN PRESS

- The proposed model is benchmarked against a wide range of traditional machine learning algorithms using a unified experimental setup to ensure fair and consistent comparisons.
- The results demonstrate that the proposed architecture achieves superior predictive performance and demonstrates promising results within the evaluated dataset, highlighting its potential applicability to real-time battery management systems.

2 RELATED WORK

DL techniques in the field of battery state estimation have shown significant improvements compared to traditional methods. Models proposed in recent years are increasingly being used in battery management systems, particularly due to their ability to model complex nonlinear degradation processes. In this section, the current literature has been classified based on the main methodological approaches in order to examine architectural trends more systematically.

2.1 Multi-Scale CNN-Based Approaches

CNNs are widely used in battery state estimation studies due to their ability to capture spatial features at different scales. Liu et al. [1] proposed a hybrid CNN–LSTM model that considers the effect of temperature changes and demonstrated that this model significantly reduces SoC estimation error. Similarly, He and Gong [4] used multi-branch CNN architectures enhanced with residual connections and attention mechanisms to improve SoH estimation accuracy under fast charging protocols.

These studies demonstrate that multi-scale CNN structures can effectively capture the spatial features that emerge under different battery operating conditions. However, approaches based solely on convolutional layers often fail to adequately represent the temporal nature of battery degradation and therefore require support from sequential modeling components.

2.2 Sequential Models Incorporating Attention Mechanisms

Attention mechanisms have become a key component in modern deep learning models by dynamically weighting important features in time series data. Qian et al. [2] proposed a CNN–SAM–LSTM architecture that simultaneously performs SoC, SoE, and SoH estimation using an uncertainty-aware loss function. Tian et al. [17], Lei et al. [19], and Hong et al. [29] demonstrated that attention mechanisms integrated into BiLSTM or GRU-based models enable more effective learning of temporal features.

Additionally, Sherkatghanad et al. [14] and Oza et al. [30] achieved more accurate SoH estimates under dynamic operating conditions using attention-enhanced CNN–BiLSTM models. Oyucu et al. [9–13] have shown that CNN–RNN combinations, LSTM, GRU, and stacking approaches provide higher accuracy than traditional regression or single machine learning methods in

ARTICLE IN PRESS

modeling the complex nonlinear degradation processes of lithium-ion batteries.

Although attention mechanisms significantly improve prediction performance, the increased model complexity and training time associated with these structures can pose practical challenges, particularly in applications with limited computational resources.

2.3 Preprocessing and Noise Reduction Strategies

Preprocessing battery data plays a critical role in improving model robustness. Chen et al. [23] used the gray relational analysis method to identify important health indicators, while Yang et al. [18] and Xie et al. [21] applied ARIMA and CEEMDAN-based methods to reduce noise in the signal and make degradation trends more apparent. Such preprocessing steps are typically used in conjunction with hybrid architectures such as CNN–LSTM or BiLSTM.

Techniques such as ARIMA, CEEMDAN, and gray relational analysis contribute to making meaningful features related to battery health more prominent by reducing signal noise. However, most of these methods are tailored to specific datasets, and generalizing them to different battery systems is not always straightforward.

2.4 Meta-Heuristic Optimization and Hybrid Approaches

In recent years, research has focused on the use of various optimization algorithms to optimize model hyperparameters and combine different model components. Zhou et al. [16] used a GA–PSO-based approach to optimize the hyperparameters of the BiLSTM–MLP–AM model, while Liang et al. [22] proposed nonlinear constraint solvers to optimize the combination of LSTM and support vector regression (SVR). In addition, Wang and Yang [31] used GAN-based data augmentation methods to reduce the limited data set problem.

Although evolutionary algorithms such as GA and PSO are effective in improving model performance, high computational costs and the limited interpretability of convergence processes can make it difficult to use these methods in real-time applications.

2.5 Graph-Based and Collaborative Learning Frameworks

Some studies have focused on graph-based or multi-output learning frameworks to model the relationships between variables in battery data. Wei and Wu [25] proposed a graph convolutional network (GCN) model with a dual attention mechanism for SOH and RUL estimation. Similarly, Xiong et al. [26] and Zhang et al. [27] developed models that combine CNN and LSTM architectures with relational structures or Kolmogorov–Arnold networks (KAN) to achieve stronger feature abstraction on missing or multi-source data.

Graph-based and multi-output models can deliver effective results in the analysis of complex or fragmented battery data. However, the high structural

ARTICLE IN PRESS

complexity of these architectures and the lack of standardized evaluation frameworks pose some challenges for practical system integration.

2.6 Dataset Usage and Benchmarking Practices

The advancement of deep learning (DL)-based battery diagnostics research has become increasingly dependent on access to higher-quality and more diverse datasets and the adoption of structured benchmarking protocols. Widely used open data sources such as the NASA Prognostics Center of Excellence (PCoE) dataset [17], the CALCE battery dataset [4], and various electric vehicle (EV) datasets [6, 26] provide important testing grounds for validating the robustness of models under realistic operating conditions. These datasets cover different battery chemistries (LFP, NMC, LCO, in this study, LCO [4]), temperature ranges, C-rate values, and driving cycles (UDDS, US06, and Highway Fuel Economy Test – HWFET). This diversity enables the development of generalizable and transferable models across different battery systems.

However, comparative evaluation practices have also become more standardized over time. Model performance is typically evaluated using metrics such as root mean square error (RMSE), mean absolute error (MAE), coefficient of determination (R^2), and mean absolute percentage error (MAPE) under different test scenarios [2, 3, 14, 16]. In addition, the consistency of preprocessing strategies is also of great importance for making fair comparisons between models. In this context, methods such as grey relational analysis [23], autoregressive integrated moving average (ARIMA) filtering [18], and CEEMDAN-based signal decomposition [21] are widely used to identify health indicators and reduce noise-induced effects prior to model training. Such procedures help ensure that performance differences between models stem from architectural features rather than differences in the dataset or preprocessing.

To enable a comprehensive comparative evaluation, this study examined 31 deep learning models published between 2020 and 2025, which are summarized in Table 1. Each model was evaluated in terms of the dataset used, architectural structure, optimization method, and reported performance results. High-performance approaches, such as CNN–BiLSTM–AM [14], ResNet–LSTM [32], and GCN models with dual attention mechanisms [25], demonstrate better generalizability when trained and validated on heterogeneous datasets containing both controlled laboratory data and real electric vehicle data.

When these studies are evaluated together, a clear trend is observed in the literature towards integrated and modular architectures that can capture the spatio-temporal degradation patterns emerging under different battery chemistries and operating conditions, moving away from singular model designs.

Between 2020 and 2025, researchers developed various deep learning architectures for battery state estimation. Figure 3 summarizes the architectural diversity in the literature by classifying 31 of these studies into

ARTICLE IN PRESS

five main categories. These categories include standard CNN–LSTM models, structures enhanced with attention mechanisms, architectures incorporating residual connections such as ResNet–LSTM, graph-based networks such as GCN, and approaches using preprocessing techniques such as DAE, ARIMA, or CEEMDAN. This classification demonstrates how architectural design preferences have evolved over time.

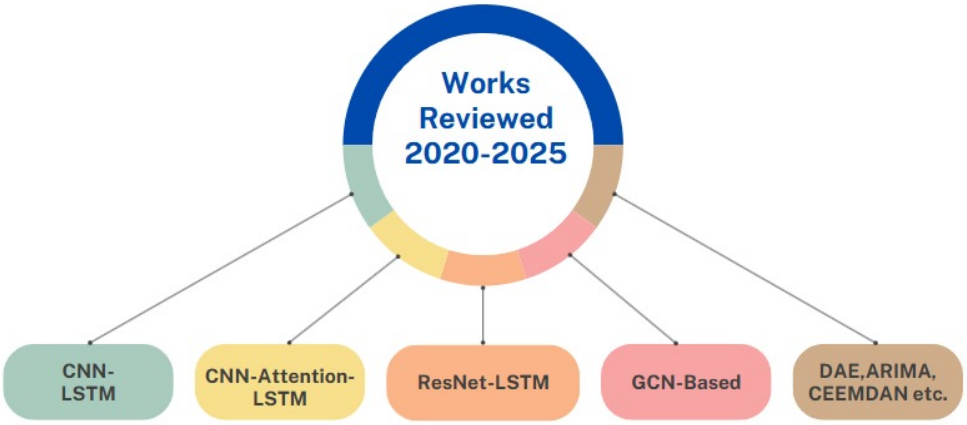


Figure 3. Classification of 31 deep learning models developed for lithium-ion battery diagnosis according to their basic architectural types. The five main categories include standard CNN–LSTM, designs enhanced with attention mechanisms, ResNet-based LSTM models, graph-based networks (GCN), and approaches with integrated preprocessing (e.g., DAE, ARIMA, CEEMDAN).

This figure illustrates the evolution of architectural approaches over time. While early studies were based on simpler, step-by-step model structures, subsequent research has seen the rise of more flexible architectures that integrate multiple components. This classification reveals that new models aim to achieve higher accuracy and stronger representation capabilities by integrating attention mechanisms, residual connections, graph-based structures, or preprocessing techniques. Furthermore, this structure provides a fundamental framework for the comparative analyses conducted in the subsequent sections of the study.

2.7 Summary of Findings from Comparative Models in Table 1

The comparative evaluation of the 31 studies listed in Table 1 reveals some significant trends in the development of deep learning-based battery state estimation models.

Dominance of Model Architectures: CNN–LSTM hybrid architectures still constitute the most widely used framework in the literature. These architectures are often enhanced with attention mechanisms [30], residual connections [32], or preprocessing modules such as ARIMA and CEEMDAN [18, 21].

Resilience Provided Through Hybridization: Hybrid temporal–spatial models (e.g., CNN + BiLSTM), when tuned with optimization methods such as GA,

ARTICLE IN PRESS

PSO, or SMA, generally perform better than single models and exhibit lower variability in RMSE and MAE values under different driving cycles and temperature conditions.

Feature Selection and Noise Reduction: Grey relational analysis [23], PCC-based filtering [30], and decomposition-based noise reduction methods such as ARIMA and CEEMDAN are widely used to highlight meaningful features in the signal and increase prediction stability.

Data Set Diversity and Generalizability: Models with higher performance are typically validated on multiple data sets. These data sets include sources such as NASA, CALCE, and real electric vehicle records; they also cover different battery chemistries (LFP, NMC, LCO) and different operating conditions such as fast charging, high C-rate values, or wide temperature ranges.

Emergence of Advanced Architectural Approaches: In recent years, new approaches such as graph-based modeling approaches (e.g., GCN) [25], quantile regression methods [15], and interval-based prediction frameworks incorporating uncertainty quantification [24] have gained attention. This trend indicates that the research community is moving towards more interpretable and flexible solutions. Furthermore, explainable AI tools such as SHAP (Shapley Additive Explanations) and LIME (Local Interpretable Model-Agnostic Explanations) are increasingly being used to enhance model reliability and diagnostic interpretability in critical applications.

In general, it is observed that the highest performance in SoC, SoH, and RUL estimation is achieved not only by increasing architectural complexity but also by the combined use of preprocessing methods, feature abstraction techniques, generalization strategies, and systematic comparative evaluation approaches.

To enable a more systematic examination of the studies summarized above, a comparative evaluation of 31 recent studies is presented in Table 1. Table entries are organized based on the study's source, purpose, data set used, proposed deep learning framework, and key findings. This structured summary highlights the fundamental trends, methodological innovations, and performance metrics emerging in data-driven methods for battery state estimation.

Table 1. Summary of DL-Based State Estimation Methods for Lithium-Ion Batteries

Authors	Purpose	Research design	Target population	Conceptual/theoretical framework	Framework proposed	Major themes
Liu et al. (2023)	To improve SoC estimation using a hybrid multi-scale CNN-LSTM approach under varying temperatures	Battery datasets under -20°C, 0°C, 25°C with MAE/RMSE analysis; experimental validation	Electric vehicle Li-ion batteries	Multi-scale convolution, LSTM memory, residual feature modeling	Multi-scale CNN-LSTM	Multi-scale CNN-LSTM architecture improves SoC prediction accuracy under dynamic temperature and driving cycles; achieves RMSE as low as 0.20% at 25°C and maintains

ARTICLE IN PRESS

						robustness across −20°C to 25°C; significant advantage over DNN, LSTM, and CNN-LSTM baselines.
Qian et al. (2024)	To develop an attention-enhanced CNN-SAM-LSTM model for estimating SOC, SOE, and SOH under realistic driving conditions	Randomized drive cycles (UDDS, BBDST, CC); split-segment evaluation; metrics: RMSE, MAE	EV battery under dynamical operating conditions	Attention-based convolution; homoscedastic uncertainty; temporal fusion	CNN-SAM-LSTM	The CNN-SAM-LSTM model simultaneously estimates SOC, SOE, and SOH under dynamic drive cycles (UDDS, BBDST, CC); achieves RMSEs below 0.9% for all parameters; introduces a homoscedastic uncertainty-aware loss function; validated across random segments and varying initial SOC values.
Zafar et al. (2024)	To accurately estimate the SoC of EV Li-ion batteries across extreme temperature conditions using a hybrid DL model optimized with the FuFi algorithm.	EV drive cycles (HPPC, HWFET, UDDS, US06); tested at -20, -10, 0, 10, 25°C; compared 5 models; metrics: RMSE, MAE, R^2, RE	EV Li-ion battery (LG 18650HG2) under temperature-stressed driving cycles	Metaheuristic optimization (FuFi), CNN for spatial feature extraction, Bi-LSTM for temporal dependency modeling	FuFi-CNN-Bi-LSTM	The FuFi-CNN-Bi-LSTM model achieves state-of-the-art SoC prediction accuracy across −20°C to 25°C; RMSE ≤ 0.00064 on US06; statistically outperforms all FuFi variants; robust under extreme temperature stress; integrates nuclear-inspired hyperparameter tuning.
He & Gong (2024)	o improve the SOH estimation of Li-ion batteries under fast-charging strategies using a hybrid neural network with a Multi-scale Convolutional Attention Mechanism (MCA)	MIT fast-charging dataset (124 Li-ion batteries); CALCE dataset ($LiCoO_2$ cathode); evaluations under limited and saturated training setups	$LiFePO_4$ and $LiCoO_2$ batteries under variable fast charging strategies	CNN, MCA (CAM+SAM), residual module; adaptive multi-branch attention with spatial-temporal enhancement	Res-MCACNN (CNN + MCA + Residual)	MCA mechanism enhances multi-scale SOH degradation feature capture; achieves lowest RMSE (0.52%) and MAE (0.37%) on CALCE dataset; shows 29–44% error reduction over SE, SK, CBAM, and hybrid baselines; adapts attention weights based on training size, improving flexibility and overfitting resistance.
Wang et al. (2023)	To improve SoC prediction by integrating a simplified electrochemical Single Particle Model (SPM) with a BiLSTM network	SPM-based internal battery modeling + BiLSTM; training with DST test data, validation with US06 and FUDS; metrics: RMSE	$LiFePO_4$ (A123 LFP 26650) battery tested under US06 and FUDS cycles	SPM for battery dynamics; residual voltage correction; BiLSTM for sequential pattern modeling	SPM-BiLSTM	Integrating SPM with BiLSTM significantly enhances SOC estimation accuracy; RMSE reduced by 33.36% over LSTM; achieves RMSE of 1.09% (US06) and 1.01% (FUDS); effectively models internal states and surface deviations using residual voltage correction.
Zhao et al. (2024)	To propose an SOH and RUL prediction model for Li-ion batteries based on AT-CNN-BiLSTM with	Experimental cycling data (voltage, temperature, current); model trained with multi-parameter time series;	Li-ion batteries under charge/discharge cycling	CNN for spatial features, BiLSTM for time-series, attention for feature weighting	AT-CNN-BiLSTM	Achieves absolute error of 0 in RUL prediction; outperforms LSTM, BiLSTM, and CNN-LSTM; demonstrates high stability and accuracy (MAPE < 0.9003, R^2 > 0.9910) for SOH and RUL predictions

ARTICLE IN PRESS

	attention mechanism	metrics: R^2 > 0.9910, MAPE < 0.9				using multi-parameter temporal data.
Authors	Purpose	Research design	Target population	Conceptual / theoretical framework	Framework proposed	Major themes
Zhang et al. (2024)	To estimate SOH using a fusion approach with Gramian Angular Field image encoding of voltage and temperature.	Battery aging experiment with voltage/temperature sensors; GAF encoding; 2D-CNN + 1D-CNN + LSTM used; metrics: RMSE, MAE	Li-ion batteries (LIBs) tested under charge/discharge conditions	Image-based SOH estimation with fusion of temperature/voltage signals using GAF and CNN	GAF + 2D-CNN + 1D-CNN + LSTM	GAF encoding improves SOH estimation by highlighting degradation features; achieves ≤2% prediction error; image-based fusion enables robust, generalizable model suitable for real-time BMS.
Bai et al. (2024)	To improve SOH prediction using a multiscale attention recurrent network (MARN) that handles long-term and short-term dynamics	Historical charge / discharge datasets with varying time scales; attention-based recurrent fusion; model validated with capacity recovery behavior	Li-ion batteries with variable SOH due to aging and cycling	Multiscale feature fusion; temporal LSTM with time-aware attention	Multiscale Attention Recurrent Network (MARN)	Effectively handles SOH estimation and capacity recovery; dual temporal scale attention improves accuracy and robustness under dynamic degradation; outperforms conventional models in non-monotonic sequences.
Sherkatghanad et al. (2024)	To propose a CNN-Bi-LSTM-AM model for SoC estimation across a wide temperature range	McMaster dataset across −20°C to 40°C; 453 training and 54K testing samples; drive cycles: US06, LA92, UDDS; metrics: RMSE, MAE	NMC Li-ion batteries (LG INR18650 HG2)	CNN for spatial, Bi-LSTM for temporal, self-attention for weight tuning	CNN-Bi-LSTM-AM	Outperforms 12 models in SoC estimation; robust across temperature conditions; RMSE ~1.8%, MAE ~1.2%; attention improves learning of complex nonlinearities and temperature variation effects.
Li et al. (2024)	To design a quantile regression-based hybrid model for accurate SoC estimation with uncertainty quantification	Data from LFP, LCO, and ternary lithium batteries; tested under various temperatures and operating conditions; metrics: MAE, RMSE	$LiFePO_4$, $LiCoO_2$, NMC batteries	TCN for fast feature extraction, GRU for temporal memory, attention for weighting; QR for interval estimation	QTGA (QR-TCN-GRU-Attention)	Combines point and interval SoC estimation with high precision (MAE < 1.45%, RMSE < 1.98%); generalizes well across battery chemistries; avoids Bayesian memory overload with QR-based loss.
Zhou et al. (2024)	To improve SOC estimation accuracy using a BiLSTM-MLP hybrid optimized by GA and PSO across multiple temperatures	Nine datasets (HPPC, DST, BBDST); tested under 5°C, 15°C, 35°C; metrics: MAE, RMSE, MAXE	EV Li-ion batteries under varying thermal conditions	GA + PSO hyperparameter optimization, BiLSTM temporal learning, MLP regression, attention mechanism	GAPSO-BiLSTM-MLP-AM	Achieves SOC estimation error <2.44% across 3 temperature ranges; demonstrates significant robustness and improved accuracy over baseline LSTM, PSO, GA; attention mechanism boosts generalization across temperatures.
Tian et al. (2022)	To predict Li-ion battery SOH by extracting health	NASA PCoE dataset; HI-based SOH prediction;	Li-ion batteries	Feature extraction of HIs, CNN for spatial	CNN-BiLSTM-AM	Extracted HI correlated to SOH; achieves RMSE < 0.01; BiLSTM captures

ARTICLE IN PRESS

	indicators and using CNN-BiLSTM-AM hybrid model	RMSE < 0.01; CNN+BiLSTM+AM trained with time series health indicators	from NASA repository	features, BiLSTM for time modeling, AM for temporal weight learning		forward/reverse dependencies while AM enhances relevance of key time features.
Yang et al. (2024)	To develop ARIMA-enhanced CNN-LSTM hybrid model for battery RUL estimation	40-cycle battery data; ARIMA applied to enhance sequence input; metrics: RMSE = 0.0136, MAE = 0.0079	Li-ion batteries under standard degradation	ARIMA for preprocessing, CNN-LSTM with attention for final estimation	ARMI-CNN-LSTM-AT	Combining ARIMA with CNN–LSTM improves battery RUL prediction accuracy. Experimental results report an RMSE of 0.0136 and demonstrate good generalization across different battery units.
Lei et al. (2024)	To build a SOC prediction model for Li-ion batteries based on SMA-optimized hybrid DL architecture	18650 battery data across 8 operation conditions; 10,596 datapoints;metrics: RMSE=0.56406% ,MAE=0.46462%, MAPE=6.80%	18650 battery packs under variable discharge conditions	CNN + Bi-GRU + attention mechanism, optimized by Slime Mold Algorithm (SMA)	SMA-CNN-BiGRU-AM	Delivers high prediction accuracy (RMSE < 0.57%); SMA tuning improves performance over non-optimized baselines; framework enhances driving range prediction and BMS integration.
Zhang et al. (2024)	To predict SOH with a denoising autoencoder-based hybrid DL model considering >10 health indicators	NASA and CALCE datasets; selected top 5 HIs by Pearson correlation; MAE = 0.0154, RMSE = 0.0191	EV batteries under real-world degradation and noise	DAE for denoising, CNN-BiLSTM for temporal-sequential modeling, attention for focus, HI filtering	DAE-CNN-BiLSTM-Attention	Effectively predicts SOH using filtered HI indicators and robust denoising; model achieves lowest errors across benchmarks; suitable for noisy, realistic data.
Authors	Purpose	Research design	Target population	Conceptual / theoretical framework	Framework proposed	Major themes
Xie et al. (2023)	To enhance RUL prediction of Li-ion batteries using CNN-LSTM with prior knowledge and CEEMDAN	NASA PCoE dataset; CEEMDAN for denoising; pretraining forms prior knowledge; evaluated on B1-B7 batteries	Li-ion batteries under experimental aging conditions	CEEMDAN for capacity sequence decomposition, CNN-LSTM for residual-based prediction, prior knowledge for weight initialization	PK-CNN-LSTM	Model reduced RMSE to 0.65%; prior knowledge enhances real-time prediction; avoids data leakage and pseudo accuracy from conventional EMD methods.
Liang et al. (2024)	To improve robustness and accuracy of SOH prediction via LSTM-SVR hybrid and nonlinear weight optimization	NASA dataset; PCA+correlation for input features; nonlinear constraint optimizer used for SVR+LSTM fusion	Li-ion batteries under standard degradation	LSTM for temporal learning, SVR for generalization, optimizer for adaptive weight allocation	Optimized LSTM + SVR	Achieved MAE < 0.75%, RMSE < 0.97%; nonlinear fusion compensates for individual model drawbacks; strong performance under noisy input
Chen et al. (2022)	To enhance RUL prediction accuracy of Li-	NASA battery data; grey relational	Li-ion batteries	Grey analysis for HI extraction, CNN for spatial	CNN-LSTM +	MAPE: 0.36%, MSE: 0.38e-4; model surpasses traditional ML and single LSTM in

ARTICLE IN PRESS

	ion batteries by combining CNN-LSTM and grey relational analysis	analysis to select key health indicators (mean current, voltage, temp)	under real usage	features, LSTM for temporal dynamics	Grey Analysis	identifying capacity decay patterns and failure thresholds
Yu et al. (2025)	To propose a BiLSTM-SDE model for interval-based SOH estimation under complex charging conditions	Open-source dataset (124 LIBs, LFP/graphite); 12 fast-charging profiles analyzed; includes transfer learning; RMSE < 0.83%	Lithium iron phosphate batteries in electric vehicles under fast charging	BiLSTM for temporal features, SDE for uncertainty modeling, transfer learning for limited-data generalization	BiLSTM-SDE-TL	Achieves interval prediction with quantified uncertainty; transfer learning allows generalization with limited data; reliable for EVs under fast and partial charging regimes.
Wei & Wu (2023)	To improve SOH and RUL predictions using graph convolutional networks with dual attention	Time-series voltage/current/temp features; graph entropy optimization for GCN structure; NASA dataset	Li-ion batteries across multiple domains	Graph entropy optimization, GCN for feature aggregation, local/global attention for temporal focus	GCN+ Dual Attention LSTM	SOH RMSE: 0.0104, RUL RMSE: 5.80; model outperforms LSTM, GPR, and boosting-based models by explicitly learning feature correlations
Xiong et al. (2024)	To develop an end-to-end SOH estimation model using incomplete multi-neighboring charging data	Real-world EV dataset (20 cars); fast-charging segments aligned by voltage; SOH from avg capacity	Electric vehicle batteries under non-uniform charging	Multi-input CNN for local feature fusion, LSTM for sequence modeling, no manual feature engineering	End-to-end Multi-Segment CNN-LSTM	Achieves MAE <2.13%, RMSE <2.74%; handles noisy and incomplete real-world EV data using aligned neighboring charging segments
Zhang et al. (2024)	To propose a KAN-LSTM hybrid for accurate SOH estimation via nonlinear function fitting and temporal modeling	Own dataset and NASA dataset; charging data under 4 C-rates; features: charge time, voltage, temperature integral	Li-ion batteries under varying C-rates and temperatures	KAN for nonlinear abstraction, LSTM for sequence learning, dropout for generalization	KAN-LSTM	Achieved MAE < 0.6%, RMSE < 0.8%, R^2 > 97%; outperformed LSTM, BiLSTM, CNN-LSTM; generalizable across own and NASA datasets with and without temperature features.
Hong et al. (2023)	To propose a dual-stage attention RNN for SOH estimation using time-dependent charging health features	NASA dataset (B5–B7, B29–B36); 4 health features extracted from charging; Spearman correlation used	Li-ion batteries in varied charging scenarios	Multi-head attention in encoder, scaled-dot product attention in decoder, LSTM backbone	Dual-Stage Attention RNN	MSE and MAPE < 0.5%; model leverages multi-head and time-dependent attention to capture rich battery aging information
Oza et al. (2024)	To develop a CNN-Attention-LSTM model for accurate and efficient SoH estimation using correlation-enhanced features	NASA battery dataset; PCC-based feature selection, normalization; model compared with existing RNN, SVR, and KF-based techniques	Li-ion batteries under variable degradation and charge conditions	CNN for spatial feature extraction, LSTM for sequential learning, attention for feature relevance tuning	CNN-Attention-LSTM	Achieves RMSE: 0.0042, MAE: 0.0039, MAPE: 0.0059; attention enhances relevant signal learning, and PCC improves feature selection efficiency; model outperforms baselines in real-time SoH prediction.
Wang &	To enhance SOH estimation via BiLSTM-WGAN	NASA B0005–B0007 (train), B0018 (test); 5	EV Li-ion batteries	WGAN-GP for data generation;	BiLSTM-WGAN	Improved SOH: MAE = 0.136, RMSE = 0.215, MAPE

ARTICLE IN PRESS

Yang (2023)	with adversarial augmentation	PCC-selected features		BiLSTM for SOH estimation		= 2.43%; outperformed LSTM-WGAN and baselines
Authors	Purpose	Research design	Target population	Conceptual/theoretical framework	Framework proposed	Major themes
Ez Eddin et al. (2024)	To present a ResNet-LSTM DL model for accurate SoH estimation in EV batteries	NASA battery dataset; evaluated across two degradation scenarios; implemented with TensorFlow	Li-ion batteries in electric vehicle settings	ResNet for residual feature extraction, LSTM for sequential SoH trends	ResNet-LSTM	Outperforms traditional ML methods (SVM, GPR, Elman NN); achieves RMSE of 0.69%; ResNet boosts deep feature learning and mitigates vanishing gradients.
Feng et al. (2023)	To estimate the SOH of Li-ion batteries under varying operating conditions using CNN-LSTM	CALCE public dataset; SOH defined by reversible capacity; tested across same/different conditions	Li-ion batteries under multi-condition scenarios	1D-CNN for feature extraction, LSTM for temporal prediction, PCC for feature selection	CNN-LSTM	The proposed model achieved an MAE of 2.33% and 3.01% in the same/different conditions, demonstrating robustness under varied working conditions; it integrates SOHC and Pearson correlation-based features.
Sun et al. (2022)	To predict the SOH of large-capacity lead–acid batteries using CNN-BiLSTM-Attention for multi-step forecasting	Time-series data from full charge–discharge cycles; model trained with temporal and spatial features for lead-acid battery; performance: <3% prediction error	Lead-acid batteries under varying load and environmental conditions	CNN for feature extraction, BiLSTM for bidirectional memory, and attention for weight adjustment	CNN-BiLSTM-Attention	Delivers accurate multi-step SOH forecasting with most errors under 3%; effective on lead–acid batteries despite structural and degradation differences from LIBs.
Alhussein et al. (2020)	To enhance individual residential electric load forecasting using a CNN-LSTM hybrid model	Smart Grid Smart City dataset (Australia); DL applied to time-series energy data; time horizons: 1, 2, 6 steps ahead	Individual residential households' electricity consumption	CNN for pattern extraction, LSTM for sequence learning, hybrid improves volatility handling	Hybrid CNN-LSTM	Improved MAPE by 4.01%–5.98% over baseline LSTM; clustering users improves accuracy; model suitable for dynamic smart grid applications.
Gao et al. (2023)	To improve RUL prediction accuracy and stability for EV Li-ion batteries using CNN-BiLSTM	NASA dataset; comparison with RNN, LSTM, BiLSTM models; evaluation based on SOH data trends	Li-ion batteries in electric vehicles	1D CNN for feature extraction, BiLSTM for bidirectional sequence learning, ELU activation	CNN-BiLSTM	The hybrid model achieved better generalization and stability than standalone models; it utilized bidirectional feature flow and deep SOH information
Zhao et al. (2021)	To propose a hybrid CNN-LSTM neural network for diagnosing battery capacity and predicting RUL	NASA PCoE dataset; experimental batteries 5, 6, 7, 33; EOL at 70%-80% capacity	Li-ion batteries under laboratory aging cycles	1D CNN for charging curve features, LSTM for long-term trend learning, ReLU activation	CNN-LSTM	Max mean RUL error: 15.65 cycles; robust end-to-end solution; overcomes manual feature extraction issues.

ARTICLE IN PRESS

As shown in Table 1, there are various developments that determine the direction of research in the field of battery state estimation. Foremost among these is the widespread use of CNN–LSTM architectures enhanced with attention mechanisms [30], uncertainty-aware loss functions [2], or physics-based preprocessing methods such as CEEMDAN [21] and ARIMA [18]. More recently, transfer learning [23], GAN-based methods for data augmentation [31], and graph-based relational modeling approaches [25] have been gaining increasing attention. This trend indicates a shift towards more generalizable and scalable model designs.

Many studies validate the robustness of their models by testing them on open-source datasets (NASA, CALCE) or specialized electric vehicle datasets. These findings show that performance gains stem not only from model complexity but also from careful feature selection, interpretability improvements, and dataset diversity. Recent studies also focus on integrating hybrid approaches that include domain adaptation and semantic consistency objectives [33], while some studies aim to develop lightweight architectures that can run on edge devices while maintaining prediction accuracy [34].

It should be noted that the studies summarized in Table 1 are conducted under different experimental conditions, including variations in datasets, preprocessing techniques, train–test splitting strategies, and evaluation metrics. Therefore, direct numerical comparison of reported performance values across these studies may not always be strictly valid. The primary purpose of this comparative analysis is to highlight general methodological trends, architectural developments, and performance ranges rather than to provide a one-to-one benchmark comparison.

2.8 Critical Analysis of Comparative Models: Strengths and Limitations

Table 1 presents a comparative summary that provides a structured overview of 31 state-of-the-art deep learning (DL) models proposed between 2020 and 2025. However, a critical and analytical evaluation requires a more in-depth examination of the methodological foundations of these models and their limitations, which are still open to debate in the literature.

Strengths: Many models in the literature demonstrate quite high prediction performance under different battery types and operating conditions. For example, the CNN–BiLSTM–attention mechanism (AM) models proposed by Sherkatghanad et al. [14] and Tian et al. [17] can achieve RMSE values below 2% thanks to attention mechanisms that more effectively weight important features in time series data. Similarly, ResNet–LSTM hybrid models [32, 35] reduce the vanishing gradient problem observed in deep networks, enabling more effective extraction of spatial features related to decay.

Furthermore, graph-based models such as GCN–Dual Attention [25] offer stronger generalization capabilities by explicitly modeling the relationships between variables. Furthermore, uncertainty-aware models such as CNN–self-attention mechanism (SAM)–LSTM [2] and BiLSTM–stochastic

ARTICLE IN PRESS

differential equation (SDE)–transfer learning (TL) [24] provide significant contributions in terms of interpretability and risk assessment. Hybrid architectures such as CNN–BiLSTM have also improved model robustness by enhancing long-term prediction performance on different battery datasets [36]. Furthermore, some studies have achieved additional performance improvements by convolutional encoding of charge curve segments followed by LSTM-based sequential modeling [37].

Furthermore, feature extraction during the preprocessing stage stands out as a significant strength. For example, gray relational analysis [23] and CEEMDAN-based decomposition [21] have been found to be highly effective in reducing noise and highlighting meaningful features in the signal. These methods particularly enhance model robustness in real-world datasets.

Limitations: Despite these developments, the models in the literature still face some significant limitations. The main ones are overfitting, scalability, and generalizability issues. For example, models trained only on NASA or CALCE datasets may not perform as well in real electric vehicle systems where factors such as temperature profiles, C-rate values, or sensor accuracy vary.

On the other hand, while models incorporating attention mechanisms offer performance advantages, these structures require a large number of hyperparameters. This situation can make it difficult to apply model compression or pruning strategies, especially in environments with limited computational resources, such as battery management systems (BMS).

Similarly, models based on meta-heuristic optimization methods such as GA–PSO [16] or the fuzzy firefly algorithm (FuFi) [3] typically have high computational costs, and the convergence behavior of the optimization process is often insufficiently transparent.

Furthermore, very few studies in the literature explicitly address model robustness against problems commonly encountered in real systems, such as missing data or sensor failures. Although explainable AI methods such as SHAP and LIME have been used in some studies for model interpretability, these approaches are mostly limited to visual explanations and do not provide a statistically robust interpretation framework.

However, recent studies have begun to take important steps toward addressing this shortcoming by using methods such as Monte Carlo-based uncertainty quantification and prediction interval generation within CNN–LSTM structures [37].

Research Gaps and Future Directions: A significant portion of the studies in the literature do not perform uncertainty quantification for prediction outputs. However, this is a critical requirement, especially in battery management systems that are critical for safety. Furthermore, the number of models using transfer learning is still quite limited [24, 26]. Transfer learning methods have the potential to increase the generalizability of models by reducing data dependency between different battery chemistries or manufacturers.

ARTICLE IN PRESS

Although comparative evaluation studies have increased, a fully standardized benchmarking framework has not yet emerged in the literature. The lack of a common standard in terms of model architectures, data set splitting strategies, performance metrics, and preprocessing protocols makes the reproducibility of studies difficult and complicates direct comparisons between models.

This study was developed to address these recurring gaps in the literature. When the 31 deep learning models examined were evaluated, three fundamental problems emerged: many models address spatial and temporal patterns separately, they can experience performance degradation under different battery types, and evaluation protocols are often inconsistent.

To address these issues, a CNN–Residual–LSTM architecture was designed, combining multi-scale spatial filtering, residual connections with the Swish activation function, and LSTM-based temporal modeling. The model structure adapts effective components from previous works such as CNN–BiLSTM–AM [14], ResNet–LSTM [32], and GCN–Dual Attention [25], and is applied on different datasets containing real electric vehicle data logs from NASA and CALCE.

The model not only produces stable predictions but also contributes to a fairer comparison of different methods. All experiments were conducted using the same performance metrics (RMSE, MAE, R^2, etc.) and the same data splitting strategies. Another frequently encountered shortcoming in the literature is the absence of uncertainty estimates in prediction outputs. Therefore, the study encourages the use of techniques such as dropout-based inference or ensemble variance to increase model reliability.

In conclusion, the proposed model does not aim to present a completely new architecture; instead, it presents a combination of components proven to be effective in the literature, validated on large datasets, and improved in terms of interpretability and robustness. This approach aims to contribute to the development of more reliable and generalizable deep learning models in battery diagnosis applications.

Whereas the comparative summary in Table 1 presents a structured overview of 31 state-of-the-art DL models put forth between 2020 and 2025, the critical reflective and analytical discussion will cut deeper into the methodological underpinnings and limitations deserving of further discourse.

Strengths: Several models exhibit outstanding predictive performance across a range of battery types and operating conditions. For instance, the CNN–BiLSTM–attention mechanism (AM) models proposed by Sherkatghanad et al. [14] and Tian et al. [17] consistently achieve sub-2% RMSE, aided by attention mechanisms that improve temporal feature relevance. Also, ResNet–LSTM hybrids [32, 35] resolve the vanishing gradient problem in deep networks so that degradation-related spatial features can be extracted effectively. Hence, Graph-based models such as GCN–Dual Attention [25] generalize well by explicitly modeling the inter-variable dependencies, while uncertainty-aware models such as CNN–self-attention mechanism (SAM)–LSTM [2] and BiLSTM–stochastic differential

ARTICLE IN PRESS

equation (SDE)–transfer learning (TL) [24] bring interpretability and risk estimation forth. In addition, hybrid architectures such as CNN–BiLSTM have demonstrated enhanced robustness and long-term prediction capability across multiple battery datasets [36], while further improvements have been achieved through convolutional encoding of charging curve segments followed by sequential LSTM layers [37].

Furthermore, feature extraction has risen as one of the strengths in pre-processing. For example, grey relational analysis [23] and CEEMDAN-based decomposition [21] have been shown to be very effective in noise reduction and maximizing signal relevance, thus ensuring higher stability of the model on real-world datasets.

Limitations: In spite of these developments, mechanisms in the literature still have to face overfitting issues, scalability, or transferability problems. For example, if a model is trained strictly on the NASA or CALCE datasets, it may not necessarily perform well on field-deployed EV systems due to differences in temperature profiles, differences in C-rates, or just differences in sensor fidelity. On another note, whilst attention-augmented solutions are welcomed, they increase the number of eligible hyperparameters to the extent that such methods cannot be readily organized given an effective time frame into strategies of compression or pruning suitable for battery management system (BMS) deployment.

In the same fashion, those optimization algorithms based heavily on metaheuristics, including, but not limited to, GA–PSO [16] and fuzzy firefly algorithm (FuFi) [3], often provide no transparency in tuning convergences while requesting large computational capabilities. Very few works explicitly address robustness to missing data or sensor faults, yet these are very common issues in real scenarios. SHAP and LIME have been mentioned for post-hoc interpretations in some works; however, their use is limited and remains more illustrative than statistically founded. Nonetheless, recent attempts have introduced MC-based quantification and prediction interval estimation within CNN–LSTM structures [37], which begin to address the statistical interpretability gap of earlier post-hoc explainability tools.

Gaps and Future Directions: A persistent limitation in many of the studies is that there is no uncertainty quantification provided on the prediction outputs, while this proves critical, especially in safety systems such as BMS. Moreover, only a very few models employ transfer learning [24, 26], which could bring down data dependency and generalizability issues across battery chemistries and manufacturers.

While decent progress has been made in benchmarking, a truly standardized framework has not yet emerged across the literature, in which model architecture templates, dataset partitions, performance metrics, and preprocessing protocols would span. This would hinder reproducibility and short-circuit model comparisons between one another.

This study was developed in response to recurring gaps observed in earlier research. After reviewing 31 DL models, three key issues became clear: many of the models treat spatial and temporal patterns separately, their

ARTICLE IN PRESS

performance drops under varying battery types, and their evaluation setups are often inconsistent. To address these points, we designed a CNN–Residual–LSTM model that brings together multiscale spatial filtering, residual pathways with Swish activation, and temporal analysis through LSTM layers. The model structure adapts effective parts from earlier designs-such as CNN–BiLSTM–AM [14], ResNet–LSTM [32], and GCN–Dual Attention [25] and applies them across a range of datasets, including NASA, CALCE, and EV logs collected under real-world conditions.

In addition to achieving stable predictions, the model improves how different methods can be compared. All experiments follow the same metric set (e.g., RMSE, MAE, R^2) and use identical data partitions to ensure fairness. Another issue found in earlier works is the absence of uncertainty estimates in the output, even though this is crucial for battery safety. This study encourages the use of techniques like dropout inference or ensemble variance to add that missing layer of reliability. The model does not aim to be a completely new architecture but rather offers a tested combination of proven elements adapted, extended, and validated for broader use in battery diagnostics where interpretability and robustness are essential.

3 MATERIAL AND METHODS

The DL architecture proposed in this paper is a hybrid model that aims to capture multiscale and contextual features of time series data. The model consists of dense layers that perform multiscale convolutions, normalization, residual connections, an LSTM layer modeling sequential dependency, and classification, respectively. Each layer is designed to maximize the capacity for in-depth feature extraction and generalization with specially specified structures. The general representation of the proposed model is given in Figure 4. The main components of the proposed model, given in Figure 4, are detailed in the subheadings.

ARTICLE IN PRESS

ARTICLE IN PRESS

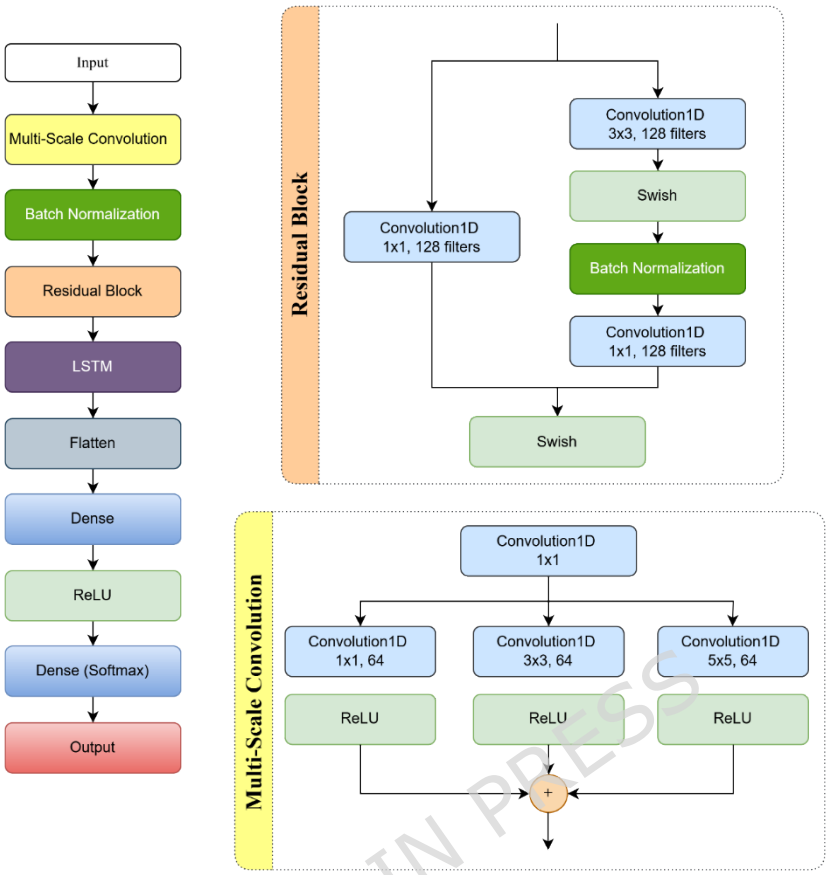


Figure 4. Proposed hybrid Multi-Scale CNN–Residual–Dense architecture for SoC estimation.

The proposed architecture illustrated in Figure 4 is designed to effectively capture multi-scale feature representations and nonlinear relationships in battery time-series data through the integration of convolutional feature extraction, residual learning, and fully connected regression layers. The main components of the proposed architecture are structured sequentially as follows:

- Step 1 (Input Representation): The model takes one-dimensional feature sequences as input, represented as (n, 1), where n denotes the number of input variables. This representation allows one-dimensional convolutional operations to be applied to extract temporal and structural patterns from the data.
- Step 2 (Multi-Scale Convolutional Feature Extraction): In this stage, a multi-branch convolutional module is employed to capture features at different receptive field sizes. Three parallel Conv1D layers are applied with kernel sizes of 1, 3, and 5, each using 128 filters and ReLU activation with same padding. The outputs of these branches are concatenated along the channel dimension to form a unified multi-scale feature representation. This structure allows the model to simultaneously learn local, mid-range, and global contextual features.

ARTICLE IN PRESS

- Step 3 (Feature Normalization): A Batch Normalization layer is applied to the concatenated feature maps to stabilize the distribution of intermediate activations and improve convergence during training.
- Step 4 (Residual Feature Refinement): An improved residual block is employed to enhance feature learning and facilitate gradient flow. The residual path consists of three convolutional layers with a kernel configuration of 1–3–1, each using 128 filters. The intermediate layers utilize the Swish activation function, and Batch Normalization is applied after each convolution. A parallel shortcut connection is implemented via a 1×1 Conv1D layer to ensure dimensional compatibility. The residual and shortcut paths are combined using element-wise addition, followed by a Swish activation, enabling effective learning of both shallow and deep representations.
- Step 5 (Feature Flattening): The refined feature maps are transformed into a one-dimensional vector using a Flatten layer, allowing the transition from convolutional representations to fully connected processing.
- Step 6 (Dense Feature Abstraction): A fully connected Dense layer with 128 neurons and ReLU activation is used to perform high-level feature abstraction and nonlinear transformation of the extracted features.
- Step 7 (Output Layer): The final prediction is obtained through a Dense layer with a single neuron and linear activation, producing the estimated State of Charge (SoC) value as a continuous output.

This step-wise architectural design enables the model to effectively integrate multi-scale feature extraction with residual learning, thereby improving representation capacity and ensuring robust performance in battery state estimation tasks.

ARTICLE IN PRESS

3.1 Critical Analysis of Benchmark Models: Strengths and Limitations

The Multi-Scale Convolutional Layer (MSCL) is designed to capture temporal patterns at different resolutions by employing parallel convolutional filters with varying kernel sizes. This structure enables the model to simultaneously learn short-, mid-, and long-range dependencies in time-series data, which is essential for modeling complex battery dynamics [38,39].

Let the input time series be defined as $T = \{t_1, t_2, \ldots, t_n\} \in R^n$. The multi-scale representation is obtained by applying multiple one-dimensional convolutional operations in parallel, each with a different kernel size.

For each scale s, the convolution operation can be expressed as:

$$Y_i^{(s)} = \phi\left(\sum_{j=1}^{k_s-1} \omega_j^{(s)} \bullet T_{i+j} + b^{(s)}\right) \quad (1)$$

ARTICLE IN PRESS

where $\omega^{(s)}$ and $b^{(s)}$ represent the filter weights and bias for the s-th scale, k_s denotes the kernel size, and $\phi(\cdot)$ is the activation function.

The outputs from different scales are then aggregated to form a unified feature representation:

$$Y = \text{Concat}(Y^{(1)}, Y^{(2)}, \ldots, Y^{(s)}) \quad (2)$$

This multi-branch design allows the model to learn features with different receptive fields without requiring explicit downsampling or additional preprocessing. Smaller kernel sizes focus on local temporal variations, while larger kernels capture broader contextual dependencies [39].

Overall, MSCL provides an efficient and scalable mechanism for multi-resolution feature extraction, significantly enhancing the model's ability to learn complex temporal patterns in battery time-series data [40].

3.2 Batch Normalization

Batch Normalization (BN) is widely used to improve training stability and accelerate convergence in deep neural networks by normalizing intermediate layer activations within each mini-batch [41]. Given a mini-batch $\{x_1, x_2, \ldots, x_m\}$, the batch mean and variance are computed as:

$$\mu = \frac{1}{m}\sum_{i=1}^{m} x_i, \sigma^2 = \frac{1}{m}\sum_{i=1}^{m} (x_i - \mu)^2 \quad (3)$$

The normalized activation is then obtained by:

$$\hat{x}_i = \frac{x_i - \mu}{\sqrt{\sigma^2 + \epsilon}} \quad (4)$$

Finally, a learnable linear transformation is applied:

$$y_i = \gamma \hat{x}_i + \beta \quad (5)$$

where γ and β are trainable scaling and shifting parameters.
Beyond normalization, BN contributes to improved optimization by smoothing the loss landscape and stabilizing gradient flow during training. As demonstrated by Santurkar et al. [42], this effect enhances convergence behavior and reduces sensitivity to initialization. From an optimization perspective, this improvement can be associated with a more controlled Lipschitz continuity of the loss function, leading to more stable gradient updates [43].

3.2.1. Impact in terms of Optimization

The analysis by Santurkar et al. revealed that the success of batch normalization (BN) is more related to making the loss function of the network smoother than reducing the "internal covariate shift". BN facilitates the optimization process by making the direction and magnitude of the gradients more stable [42].

ARTICLE IN PRESS

This effect is expressed in terms of Lipschitz continuity as follows:

$$|L(x_i) - L(x_2)| \leq L\|x_1 - x_2\|, \forall x_1, x_2 \quad (6)$$

This inequality shows that BN improves learning dynamics by limiting the local slope of the loss function [43].

3.3 Residual Block

As network depth increases, training deep neural networks becomes more challenging due to issues such as vanishing gradients and degraded performance. Residual blocks address this problem by learning a residual mapping instead of directly approximating the desired function [44]. This formulation can be expressed as:

$$y = F(x,W) + x \quad (7)$$

where x is the input and $F(x,W)$ represents the residual function parameterized by weights W. The skip connection enables direct gradient flow, facilitating stable training in deeper architectures [45]. In this study, the residual block is designed using a lightweight convolutional structure. The residual mapping is defined as:

$$F(x) = \text{Conv1D}_{3\times3} \rightarrow \text{Swish} \rightarrow \text{BN} \rightarrow \text{Conv1D}_{1\times1} \quad (8)$$

To ensure dimensional compatibility, the shortcut path is transformed as:

$$x' = \text{Conv1D}_{1\times1}(x) \quad (9)$$

The final output of the residual block is obtained as:

$$y = \text{Swish}(F(x) + x') \quad (10)$$

This design enables the model to effectively combine low-level and high-level feature representations while maintaining efficient gradient propagation. As a result, the residual structure improves both learning stability and representation capacity in deep architectures [46,47].

3.4 The Long Short-Term Memory

Long Short-Term Memory (LSTM) networks were specifically designed to address the vanishing gradient problem encountered in conventional Recurrent Neural Networks (RNNs), thereby enabling the effective modeling of long-term dependencies in sequential data [48,49]. Unlike standard RNNs, LSTM introduces a memory cell structure that can preserve information over extended time intervals, making it particularly suitable for time-series modeling tasks such as battery state estimation.

The core functionality of LSTM relies on a set of gating mechanisms—namely, the forget gate, input gate, and output gate—which regulate the flow of information into, within, and out of the memory cell [50]. These gates allow the network to selectively retain relevant historical information, incorporate

ARTICLE IN PRESS

new inputs, and generate context-aware outputs, thus enhancing both learning stability and representational capacity.

At each time step t, the LSTM cell processes the current input x_t, the previous hidden state h_{t-1}, and the previous cell state C_{t-1}. The internal operations of the LSTM unit can be expressed in a unified mathematical formulation as follows [51,52]:

$$f_t = \sigma(W_f.[h_{t-1},x_t] + b_f) \quad (11)$$

$$i_t = \sigma(W_i.[h_{t-1},x_t] + b_i) \quad (12)$$

$$\tilde{C}_t = \tanh(W_C.[h_{t-1},x_t] + b_C)\,, \quad (13)$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \quad (14)$$

$$O_t = \sigma(W_o.[h_{t-1},x_t] + b_o) \quad (15)$$

$$h_t = o_t \odot \tanh(C_t) \quad (16)$$

In these equations, σ(·) denotes the sigmoid activation function, tanh(·) represents the hyperbolic tangent function, and ⊙ indicates element-wise multiplication. The forget gate f_t controls the retention of previous memory, while the input gate i_t and candidate state $\hat{C}_t$ determine how new information is incorporated. The updated cell state C_t serves as the internal memory of the network. Finally, the output gate oto_tot regulates the information passed to the hidden state hth_tht, which is used for subsequent predictions.

Through this gated structure, LSTM networks can effectively capture both short-term variations and long-term dependencies in sequential data, making them particularly suitable for modeling the nonlinear temporal dynamics observed in lithium-ion battery systems.

ARTICLE IN PRESS

3.5 Flatten and Dense Layers

In convolutional neural networks, the multi-dimensional feature maps obtained from convolutional layers are transformed into a one-dimensional representation through a flattening operation [53]. Let $X \in R^{d_1 \times d_2 \times \cdots \times d_k}$ denote the input tensor; the flattening process reshapes it into a vector $x \in R^n$, where:

$$n = \prod_{i=1}^{k} d_i \quad (17)$$

This vector representation enables the transition from spatial feature extraction to fully connected processing. The flattened vector is then fed into

ARTICLE IN PRESS

fully connected (Dense) layers, which perform a linear transformation followed by a nonlinear activation:

$$y = f(Wx + b) \quad (18)$$

where $W \in R^{m \times n}$ and $b \in R^{m}$ denote the weight matrix and bias vector, respectively, and $f(\cdot)$ is an activation function, typically the Rectified Linear Unit (ReLU). Dense layers facilitate high-level feature abstraction by integrating information across all input dimensions. However, the flattening operation may significantly increase the number of trainable parameters, which can lead to higher computational cost and an increased risk of overfitting if not properly regularized.

3.6 Activation Functions

Activation functions introduce nonlinearity into deep neural networks, enabling the modeling of complex patterns in data. In this study, both Swish and ReLU activation functions are utilized due to their complementary properties.

The Swish activation function is defined as:

$$\text{Swish}(x) = x \bullet \sigma(\beta x) \quad (19)$$

where $\sigma(\bullet)$ denotes the sigmoid function and β is a fixed or learnable parameter [54]. Swish is smooth and continuously differentiable, which facilitates stable gradient flow during backpropagation. Unlike ReLU, it allows small negative values to pass through, thereby reducing the risk of inactive ("dead") neurons and enabling richer feature representations [55,56].

The Rectified Linear Unit (ReLU), defined as:

$$\text{ReLU}(x) = \max(0,x) \quad (20)$$

is widely used due to its simplicity and computational efficiency [57–59]. It effectively mitigates the vanishing gradient problem in deep networks by preserving positive activations, although it may suffer from zero-gradient regions for negative inputs.

In the proposed model, Swish is primarily employed within residual blocks to enhance feature learning through smoother optimization dynamics, while ReLU is used in fully connected layers to maintain computational efficiency. This combined usage leverages the advantages of both activation functions, contributing to improved model stability and generalization performance [60].

4 EXPERIMENTAL RESULTS

The experiments were conducted on a system with an Intel Core i7-12700H processor, NVIDIA GeForce RTX 3070 Ti GPU, and 32 GB of RAM.

ARTICLE IN PRESS

4.1 Performance Metrics

Three basic regression metrics were used to assess the accuracy of the model: MAE, MSE, and R^2. These metrics are defined by the following mathematical expressions:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right| \quad (21)$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (22)$$

$$R - Squared = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (23)$$

Where y_i is the true values; $\hat{y}_i$ is the values predicted by the model; $\bar{y}$ is the mean of the true values; and n is the number of samples. These three metrics provide a holistic analysis to assess not only the level of error but also the stability and generalizability of the model. The outputs of the model are calculated and reported separately on both training and test data according to these metrics.

4.2 Implementation Details

Before training the model, all input features were normalized to the range [0, 1] using the MinMaxScaler. The normalization parameters were computed from the training set only and then applied to the test set to prevent data leakage. Using a random hold-out strategy, the dataset was divided into training and test subsets with an 80%-20% split. During training, 20% of the training data was reserved as a validation set to monitor the model's generalization performance. This validation split was performed once and used consistently across all experimental runs to ensure comparability. Although the dataset was randomly partitioned, the original sample indices were preserved and used during the evaluation phase. Specifically, the test samples were reordered according to their original temporal/sample index when generating SOC estimation curves and related visualizations. This approach allows for a realistic, time-indexed interpretation of the model predictions while maintaining the benefits of random data splitting. Consequently, the results were ensured to be both statistically reliable and temporally interpretable.

Two basic early stopping strategies were used to reduce the risk of overfitting during training. With the EarlyStopping method, training was terminated when no improvement in verification loss was observed for 75 consecutive epochs, and with the ReduceLROnPlateau mechanism, the learning rate was halved if the verification performance did not improve for 15 epochs. In this way, the model was stopped in case of early saturation, and more precise optimization was achieved.

For the optimization of the model, the Adam algorithm was preferred, and MSE was used as the loss function. Although the maximum training time was set to 2000 epochs, training was usually completed earlier. After the training

ARTICLE IN PRESS

was completed, predictions were generated on the test set using the best weights, and the final performance metrics were calculated based on these predictions.

It should be noted that the proposed model is designed to produce deterministic predictions, and explicit uncertainty quantification mechanisms are not incorporated within the scope of the current implementation. While the achieved performance metrics indicate high predictive accuracy, they do not directly provide information about the confidence of individual predictions. Incorporating uncertainty-aware approaches, such as Monte Carlo dropout or ensemble-based methods, may further enhance the reliability of the model, particularly for safety-critical battery management applications, and is considered as a direction for future work.

4.3 Results

In this section, the performance of the developed hybrid DL architecture on SoC estimation is systematically evaluated in the light of the contributions of the different components used in the architecture. The proposed model captures local spatial features with multi-scale CNN, enhances the learning capability of deep structures with improved residual blocks, and models time-series specific sequential dependencies with LSTM layers. In order to analyze the impact of each of these components on model performance, different combinations were tried, and each architecture was trained with the same dataset, the same training/test split, and fixed hyperparameter settings. The results are summarized in Table 2.

Table 2. Effect of different model components on SoC prediction performance

Multi_scale_cnn	Residual_block	LSTM	Performance Metrics		
			MAE	MSE	R^2
x			0.2891	0.3870	0.9995
	x		0.3377	0.5297	0.9994
		x	0.4870	0.8827	0.9989
x	x		0.2622	0.3712	0.9996
x		x	0.1869	0.2555	0.9997
x	x	x	0.1507	0.2030	0.9998

Examining Table 2 reveals that the model components produce high accuracy values when used separately. However, the proposed hybrid structure that integrates these components significantly improves performance in all metrics. In the model using only the LSTM layer, for

ARTICLE IN PRESS

example, the MAE value is 0.4870, the MSE value is 0.8827, and the R^2 score is 0.9989. This indicates that the sequential structure alone cannot provide sufficient generalization capability. Conversely, in the model where only convolutional structures were used, specifically the multi-scale CNN module, error rates decreased (MAE: 0.2891; MSE: 0.3870); however, accuracy remained limited since temporal dependencies were not explicitly modeled.

The model using a convolutional structure and LSTM together reduced the MAE to 0.1869 and the MSE to 0.2555, demonstrating the effectiveness of combining temporal and spatial information. However, the proposed hybrid architecture with all three components achieved the lowest error values and the highest R^2 score (MAE: 0.1507, MSE: 0.2030, R^2: 0.9998). These results demonstrate that the model achieves a high level of predictive accuracy in both point estimation and variance-explaining ability.

The R^2 score of 0.9998 indicates that the model can explain 99.98% of the variance in state of charge (SoC) and offers high accuracy, low error, and strong generalization capacity under the evaluated experimental setting. These results demonstrate the strong potential of the proposed model for accurately predicting SoC in energy management and battery monitoring applications.

In addition to the quantitative findings, Figure 5 presents a scatter plot of the actual and predicted SoC values to visually demonstrate the model's prediction accuracy. The proximity of the points to the 45-degree reference line ($y = x$) in the graph clearly shows the accuracy of the model's predictions. The 45-degree reference line represents ideal prediction performance. The majority of the points align extremely close to the reference line, revealing that the model can make consistent predictions without producing any systematic bias. Notably, the predictions do not significantly deviate from the baseline, even at extreme values, demonstrating the model's ability to perform consistently and generalize to extreme cases. This graphical validation corroborates the model's consistent performance, consistent with the high R^2 score and low error metrics quantified in previous sections.

ARTICLE IN PRESS

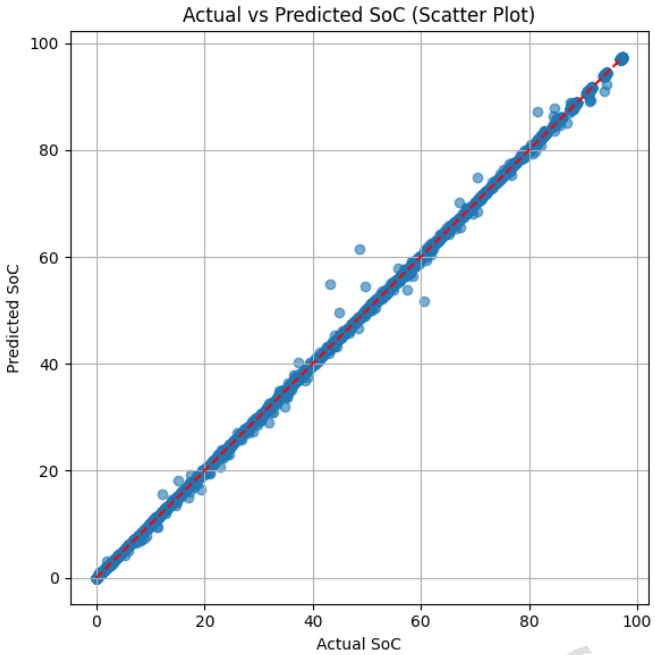


Figure 5. Relationship between model-predicted SoC values and actual SoC values

Additionally, Figure 6 visualizes the distribution of residuals to further evaluate the model's prediction performance. This graph shows the differences between the predicted state of charge (SoC) value and the actual value for each test case on a sequential sample basis. The residuals are distributed primarily around the zero baseline, suggesting that the model produces predictions without a clear systematic bias. Additionally, the small number of outliers and their irregular distribution suggest that the model avoids strong overfitting behavior. These findings support the idea that the model has a stable error distribution across the evaluated test samples.

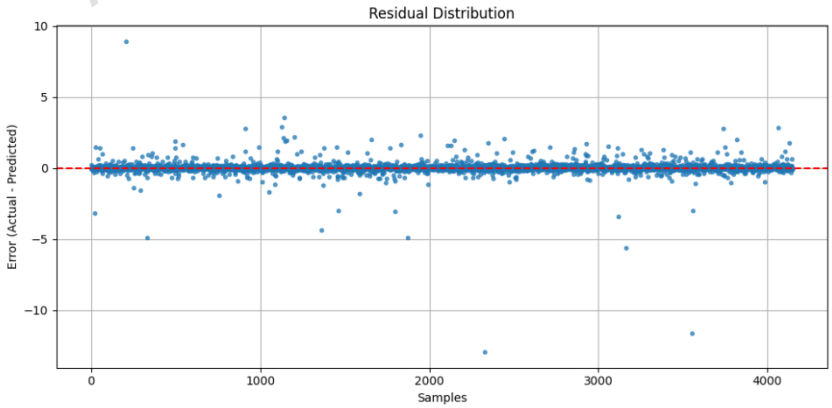


Figure 6. Sample-based distribution of the model's prediction errors.

To provide a more comprehensive evaluation of model behavior, we present SOC estimation curves based on the original sample index. Figure 7 shows the time-indexed state of charge (SOC) estimation results for multiple model

ARTICLE IN PRESS

configurations, including the proposed model and baseline architectures. The curves reflect the order of the actual test samples, allowing for a realistic interpretation of the model's ability to track dynamic changes in battery behavior. Compared to other models, the proposed CNN–Residual–LSTM model demonstrates closer alignment with the actual SOC values across the entire sequence.

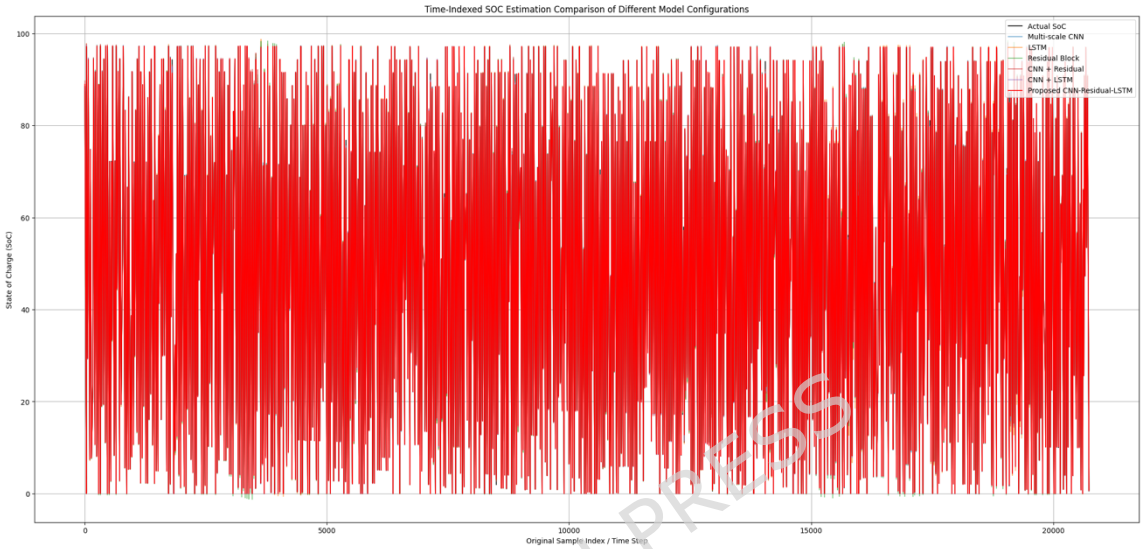


Figure 7. Time-indexed SOC estimation comparison of different model configurations.

To improve interpretability in regions with dense fluctuations, Figure 8 provides a zoomed-in view. In this figure, the ground truth state of charge (SOC) is highlighted in black, and the proposed model is emphasized in red. This visualization clearly shows that the proposed model has higher prediction stability and accuracy, particularly in segments with rapid SOC changes.

ARTICLE IN PRESS

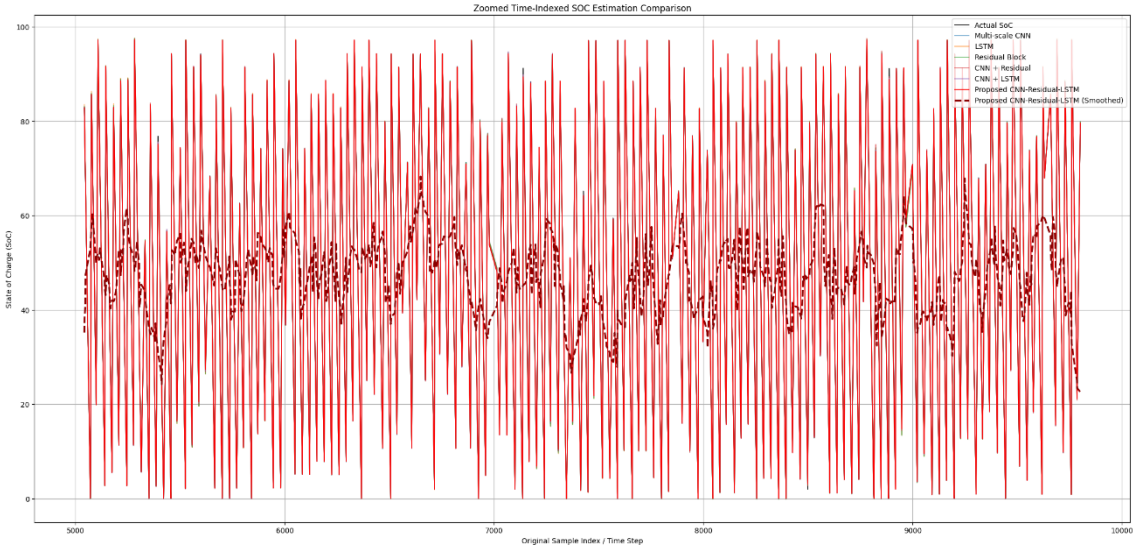


Figure 8. Zoomed time-indexed SOC estimation comparison for selected intervals.

Furthermore, the models' prediction performance is analyzed in terms of absolute error across the time axis. Figure 9 shows the absolute error comparison of different model configurations over time. The results show that the proposed model generally produces lower error values than the baseline configurations, especially in regions with high variability.

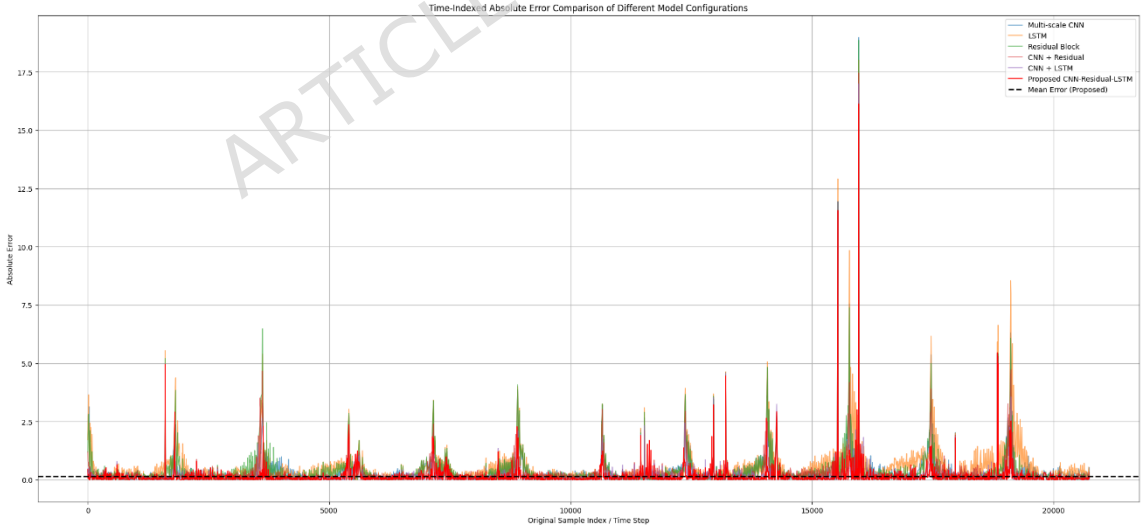


Figure 9. Time-indexed absolute error comparison of different model configurations.

Additionally, Figure 10 provides a logarithmic-scale representation of the absolute error, allowing for a more detailed analysis across different error magnitudes. This visualization allows you to examine small and large

ARTICLE IN PRESS

deviations simultaneously, providing additional insight into the error characteristics of the compared models.

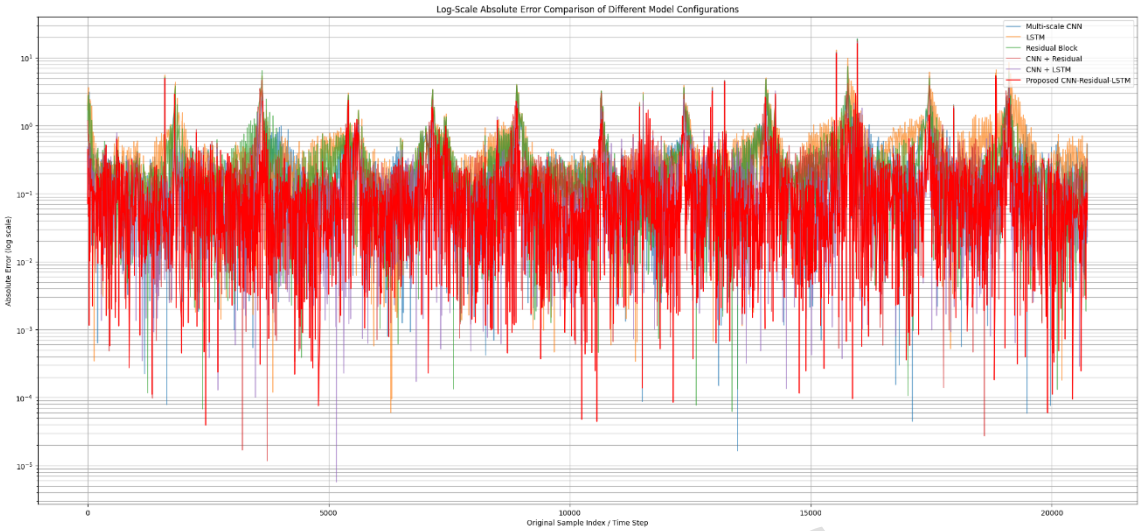


Figure 10. Log-scale absolute error comparison of different model configurations.

Overall, quantitative metrics and visual analyses consistently demonstrate that integrating multi-scale convolutional feature extraction, residual learning, and long short-term memory (LSTM)-based temporal modeling enhances SoC estimation performance. The results suggest that the proposed hybrid architecture offers a more robust and precise prediction framework than the individual component-based configurations examined in this study.

5 DISCUSSION

This study analyzes the performance of the hybrid DL model in comparison with various traditional ML algorithms. Despite the different model structures, the comparison process is rigorously structured to ensure a fair and direct evaluation. In this context, all algorithms were trained and tested on the same dataset with an 80:20 training-to-testing ratio and a fixed random_state parameter. This avoids random deviations due to data fragmentation and increases the validity of the comparisons.

In the data preprocessing process, only numeric variables were considered, and all features were scaled to the range [0, 1] with MinMaxScaler. The normalization process was performed only on the training data, and the same transformation parameters were applied to the test data to avoid data leakage. This approach ensures consistency in comparative analysis between DL architecture and traditional models. All traditional ML algorithms used in the comparative analysis were implemented through the scikit-learn library,

ARTICLE IN PRESS

and the models were trained with the default parameter values of the respective classes. All models were evaluated using three main performance metrics on the test data: MAE, MSE, and R^2. The obtained performance values are presented in Table 3.

Table 3. Comparative performance of traditional methods and the proposed model

		MAE	MSE	R-Squared
Traditional ML algorithms	Linear Regression (LR)	4.1831	26.4805	0.9689
	AdaBoost (AB)	2.4498	9.2155	0.9892
	SVR	1.0606	2.3520	0.9972
	GBM	0.4690	0.6953	0.9991
	Extra Trees (ET)	0.2610	0.7755	0.9990
	Bagging (BG)	0.2476	0.7190	0.9991
	Decision Tree (DT)	0.2431	1.0818	0.9987
	RF	0.2311	0.6732	0.9992
	K-Nearest Neighbors (KNN)	0.2222	0.7387	0.9991
Deep Learning algorithms	LSTM	0.4870	0.8827	0.9989
	GRU	0.4983	0.8910	0.9989
	1DCNN	0.4293	0.8067	0.9990
	Önerilen Model	0.1507	0.2030	0.9998

As shown in Table 3, the comparative results demonstrate that the proposed hybrid DL model outperforms all traditional ML algorithms. In particular, the low MAE and MSE values indicate that the model's prediction error is quite low. Regarding MAE, the proposed model produced a better result than its closest competitor, the RF Regressor, with an error that is about 34% lower (0.1507 vs. 0.2311). Similarly, the proposed model had about 70% lower MSE, with a value of 0.2030, compared to RF's value of 0.6732. The R^2 score of 0.9998 is very close to 1, indicating a very high explanatory capability of the model in capturing the variance of the SoC values. This value exceeds even the R^2 score of RF, the most successful traditional method, indicating the model's high overall learning capacity.

In addition to traditional machine learning (ML) models, the comparative analysis included deep learning-based baseline architectures, such as long short-term memory (LSTM), gated recurrent unit (GRU), and one-dimensional convolutional neural network (1D CNN), to provide a more comprehensive experimental validation. As shown in Table 3, these models perform relatively well; however, they remain significantly behind the proposed hybrid architecture in all evaluation metrics. For example, the standalone LSTM model yields an MAE of 0.4870 and an MSE of 0.8827,

ARTICLE IN PRESS

while the GRU model produces similar results (MAE: 0.4983, MSE: 0.8910). This indicates that modeling sequences alone is insufficient to capture the full complexity of SoC dynamics. Similarly, the 1D CNN model, which focuses on spatial feature extraction, outperforms recurrent models but still falls short of the proposed architecture (MAE: 0.4293, MSE: 0.8067). These results clearly show that neither purely temporal nor purely spatial models can individually achieve optimal performance.

The superior performance of the proposed model is due to its hybrid architecture, which effectively integrates multi-scale convolutional feature extraction, residual learning, and temporal sequence modeling. This combination allows the model to capture short-term fluctuations and long-term dependencies in battery behavior, significantly improving prediction accuracy. The consistent improvement observed across all evaluation metrics confirms that the proposed architecture effectively addresses the limitations of both traditional ML approaches and standalone deep learning models. Furthermore, ensemble-based models (RF, BG, GBM, etc.) also remain inferior to the proposed architecture, indicating that the performance gain is not solely due to model complexity but rather the effective integration of complementary learning mechanisms.

The comparisons show that the proposed model can more successfully learn non-linear relationships and interactions while reducing the overall error level. The model demonstrates better overall performance than traditional methods, especially in complex, multidimensional systems. Additionally, the balance between error metrics indicates that the model has a robust structure in terms of general accuracy and generalization capability. Despite these promising results, the reliability of the model must be considered for safety-critical applications such as battery management systems. The proposed architecture in the current study produces deterministic point estimates without explicitly quantifying predictive uncertainty. While the high accuracy achieved (low MAE and MSE, and high R^2) indicates strong predictive capability, these metrics do not directly provide information about the confidence or robustness of individual predictions.

Recent studies in battery diagnostics underscore the importance of uncertainty-aware modeling approaches in enhancing reliability and facilitating risk-sensitive decision-making. Techniques such as Monte Carlo dropout, deep ensemble learning, and Bayesian neural networks enable us to estimate predictive distributions and confidence intervals. These provide additional insight into how models behave under different operating conditions. However, the primary objective of this study was to develop a high-performance hybrid architecture that integrates multiscale convolutional feature extraction and residual learning to accurately estimate

ARTICLE IN PRESS

state of charge (SoC). Therefore, uncertainty quantification was beyond the scope of this work. Nevertheless, integrating such mechanisms into the proposed framework is an important direction for future research, particularly to improve model trustworthiness and enable deployment in real-time battery management systems.

6 CONCLUSION

In this study, a novel hybrid DL architecture using multi-scale CNNs, enhanced residual blocks and LSTM units is proposed to predict the SoC value of battery systems with high accuracy. The proposed model effectively learns spatial patterns and temporal dependencies in the data and achieves accurate predictions in the regression task.

This study's primary contribution is the design of a unified hybrid architecture that integrates multi-scale convolutional feature extraction (with kernel sizes of 1, 3, and 5), Swish-activated residual learning, and temporal modeling within a single framework. Unlike conventional CNN–LSTM and ResNet–LSTM variants, this approach more effectively represents local and global feature patterns, improving learning capability and stability. Additionally, controlled ablation analysis and a consistent benchmarking strategy were employed to ensure fair and interpretable evaluation of the model components.

The experimental results indicate that the proposed model achieves higher prediction accuracy than traditional ML algorithms (RF, GBM, SVR etc.) and performs better than simplified versions of the model with reduced components. The R^2 score of the model is 0.9998, the MAE value is 0.1507, and the MSE value is 0.2030, indicating that the prediction performance is both highly accurate and has strong generalization capacity. Moreover, the scatter plot and residual distribution analyses visually confirm that the model offers a balanced and consistent forecasting capability that is free from systematic errors. In this context, the proposed hybrid architecture provides accurate predictions and demonstrates stable performance across the evaluated datasets for real-time SoC prediction and offers the potential as a viable decision support tool for smart battery management systems.

In addition, from a computational perspective, the proposed architecture is based on 1D convolutional operations and a relatively simple residual structure, which may offer advantages in terms of computational efficiency compared to more complex architectures involving high-dimensional convolutions or attention-heavy mechanisms. While a detailed quantitative analysis of computational cost, inference latency, and memory requirements is beyond the scope of this study, these aspects will be investigated in future work to further assess the deployment feasibility of the proposed model in resource-constrained battery management system environments.

Despite these strong results, it is important to note that the proposed model makes deterministic predictions and does not quantify predictive uncertainty. While the performance metrics achieved indicate high accuracy, they do not

ARTICLE IN PRESS

directly reflect the confidence level associated with individual predictions. Given the safety-critical nature of battery management systems, incorporating uncertainty-aware modeling approaches would enhance the reliability and practical applicability of the proposed framework.

Future studies will focus on adapting the model to online learning mechanisms, integrating it with hardware-level battery monitoring units, and incorporating uncertainty quantification techniques, such as dropout-based inference and ensemble learning, to enable confidence-aware and risk-sensitive decision-making. Additionally, testing the model under various degradation scenarios and real-world operating conditions will improve its robustness and readiness for deployment.

Funding Declaration

The authors declare that no financial support, grants, or funding were received for the conduct of this study.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions Statement

Conceptualization, S.O. and E.B.; methodology, S.O.; validation, A.A. and E.B.; formal analysis, E.B. and F.K.; investigation, E.B. and S.O.; resources, A.A.; data curation, O.P., M.T., E.B.; writing—original draft preparation, S.O., E.B., O.P., M.T., F.K., A.A.; writing—review and editing, S.O. and E.B.; supervision, E.B. All authors have read and agreed to the published version of the manuscript.

Acknowledgments

The authors gratefully acknowledge the resources and infrastructure provided by Sivas University of Science and Technology and Gazi University, which significantly contributed to the successful completion of this research.

Data availability

The datasets generated during the current study are available in the Github repository at

https://github.com/mturkoglu23/Hybrid_Multi-Scale_CNN

REFERENCES

[1] Liu, C., Wang, B. & Zou, Y. Combining multi-scale convolutional neural network with long short-term memory neural network for state of charge estimation of lithium-ion batteries. *Proc. 2nd Int. Conf. Robotics, Automation and Intelligent Control (ICRAIC 2022)*, Vol. 2456, 012017 (2023). https://doi.org/10.1088/1742-6596/2456/1/012017

[2] Qian, C., *et al.* A CNN-SAM-LSTM hybrid neural network for multi-state estimation of lithium-ion batteries under dynamical operating conditions. *Energy* **294**, 130764 (2024). https://doi.org/10.1016/j.energy.2024.130764

[3] Zafar, M. H. *et al.* 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 (2024). https://doi.org/10.1016/j.energy.2024.130584

ARTICLE IN PRESS

[4] Diao, W., Saxena, S. and Pecht, M. Accelerated cycle life testing and capacity degradation modeling of $LiCoO_2$-graphite cells. *J. Power Sources* **435**, 2019, 226830 (2019). https://doi.org/10.1016/j.jpowsour.2019.226830
[5] Wang, F. *et al.* Lithium-ion battery SOC estimation by integrating electrochemical model with deep learning. *Proc. 2023 3rd New Energy and Energy Storage System Control Summit Forum (NEESSC)* (2023). https://doi.org/10.1109/NEESSC59976.2023.10349281
[6] Zhao, F. M. Gao, D. X., Cheng, Y. M., and Yang, Q. Application of state of health estimation and remaining useful life prediction for lithium-ion batteries based on AT-CNN-BiLSTM. *Sci. Rep.* **14**, 29026 (2024). https://doi.org/10.1038/s41598-024-80421-2
[7] Zhang, W. *et al.* Lithium-ion battery state of health prognostication employing multi-model fusion approach based on image coding of charging voltage and temperature data. *Energy* **296**, 131095 (2024). https://doi.org/10.1016/j.energy.2024.131095
[8] Bai, X. *et al.* Multiscale attention recurrent network for storage estimation of lithium-ion batteries. *Proc. 2024 43rd Chinese Control Conference (CCC)*, 8399–8404 (2024). https://doi.org/10.23919/CCC63176.2024.10662788
[9] Oyucu, S., Doğan, F., Aksöz, A. & Biçer, E. Comparative analysis of commonly used machine learning approaches for Li-ion battery performance prediction and management in electric vehicles. *Appl. Sci.* **14**, 2306 (2024). https://doi.org/10.3390/app14062306
[10] Oyucu, S., Dümen, S., Duru, İ., Aksöz, A., Biçer, E. Discharge capacity estimation for Li-ion batteries: A comparative study. *Symmetry* **16**, 436 (2024). https://doi.org/10.3390/sym16040436
[11] Oyucu, S., Ersöz, B., Sağıroğlu, Ş., Aksöz, A., Biçer, E. Optimizing lithium-ion battery performance: Integrating machine learning and explainable AI for enhanced energy management. *Sustainability* **16**, 4755 (2024). https://doi.org/10.3390/su16114755
[12] Çetinus, B., Oyucu, S., Aksöz, A., Biçer, E. The role of machine learning in enhancing battery management for drone operations: A focus on SoH prediction using ensemble learning techniques. *Batteries* **10**, 371 (2024). https://doi.org/10.3390/batteries10100371
[13] Ersöz, B., Oyucu, S., Aksöz, A., Sağıroğlu, Ş., Biçer, E. Interpreting CNN-RNN hybrid model-based ensemble learning with explainable artificial intelligence to predict the performance of Li-ion batteries in drone flights. *Appl. Sci.* **14**, 10816 (2024). https://doi.org/10.3390/app142310816
[14] Sherkatghanad, Z., Ghazanfari, A. & Makarenkov, V. A self-attention-based CNN-Bi-LSTM model for accurate state-of-charge estimation of lithium-ion batteries. *J. Energy Storage* **88**, 111524 (2024). https://doi.org/10.1016/j.est.2024.111524
[15] Li, H., Fu, L., Long, X., Liu, L., and Zeng, Z. A hybrid deep learning model for lithium-ion batteries' state of charge estimation based on quantile regression and attention. *Energy* **294**, 130834 (2024). https://doi.org/10.1016/j.energy.2024.130834

ARTICLE IN PRESS

[16] Zhou, Y., Wang, S., Zeng, J., Xie, Y., Fernandez, C. SOC estimation of lithium-ion battery based on optimized bidirectional long short term memory–multilayer perceptron hybrid model at multi-temperature ranges. *SSRN* (2024). https://doi.org/10.2139/ssrn.4695717
[17] Tian, Y., Wen, J., Yang, Y., Shi, Y., and Zeng, J. State-of-health prediction of lithium-ion batteries based on CNN-BiLSTM-AM. *Batteries* **8**, 155 (2022). https://doi.org/10.3390/batteries8100155
[18] Yang, L., Hanqing, Z. & Lingfeng, W. ARIMA and attention-based CNN-LSTM hybrid neural network for battery life estimation. *Proc. 2024 8th Int. Conf. Machine Learning and Soft Computing (ICMLSC 2024)* (2024). https://doi.org/10.1145/3647750.3647759
[19] Lei, W., Gu, X. & Zhou, L. Deep learning model based on the introduction of attention mechanism for lithium-ion battery SOC prediction. *J. Electrochem. Soc.* **171**, 074509 (2024). https://doi.org/10.1149/1945-7111/ad5d1e
[20] Zhang, C., Hu, Y., Huang, D., Fang, J., and Qin, N. A data-driven DAE-CNN-BiLSTM-attention prediction model for state of health of lithium-ion batteries. *Proc. 2024 14th Int. Conf. Information Science and Technology (ICIST)*, 155–164 (2024). https://doi.org/10.1109/ICIST63249.2024.10805337
[21] Xie, Q., Liu, R., Huang, J., Dong, L., and Su, J. Residual life prediction of lithium-ion batteries based on data preprocessing and a priori knowledge-assisted CNN-LSTM. *Proc. 2023 IEEE 6th Int. Electrical and Energy Conf. (CIEEC)*, 1128–1133 (2023). https://doi.org/10.1109/CIEEC58067.2023.10166412
[22] Liang, Y. *et al.* State of health prediction of lithium-ion batteries using combined machine learning model based on nonlinear constraint optimization. *J. Electrochem. Soc.* **171**, ad18e1 (2024). https://doi.org/10.1149/1945-7111/ad18e1
[23] Chen, D., Zheng, X., Chen, C., and Zhao, W. Remaining useful life prediction of lithium-ion battery based on CNN-LSTM fusion model and grey relational analysis. *Electron. Res. Arch.* **31**, 633–655 (2022). https://doi.org/10.3934/era.2023031
[24] Yu, X., Tang, T., Song, Z., and He, Y. State-of-health estimation for lithium-ion batteries under complex charging conditions based on SDE-BiLSTM model. *J. Energy Storage* **111**, 115352 (2025). https://doi.org/10.1016/j.est.2025.115352
[25] Wei, U. & Wu, D. Prediction of state of health and remaining useful life of lithium-ion battery using graph convolutional network with dual attention mechanisms. *Reliab. Eng. Syst. Saf.* **230**, 108947 (2023). https://doi.org/10.1016/j.ress.2022.108947
[26] Xiong, X. *et al.* End-to-end deep learning powered battery state of health estimation considering multi-neighboring incomplete charging data. *Energy* **292**, 130495 (2024). https://doi.org/10.1016/j.energy.2024.130495

ARTICLE IN PRESS

[27] Zhang, Z. *et al.* Advanced state-of-health estimation for lithium-ion batteries using multi-feature fusion and KAN-LSTM hybrid model. *Batteries* **10**, 433 (2024). https://doi.org/10.3390/batteries10120433
[28] Fu, J., Song, Z., Meng, J., Guo, J., Yang, K., Liu, W., & Huan, L. AI-augmented electrochemical model for lithium-ion battery: Recent advances and perspectives. Journal of Energy Chemistry, 113 (2026), 1056–1080. https://doi.org/10.1016/j.jechem.2025.10.008
[29] Hong, J., Chen, Y., Chai, Q., Lin, Q., Wang, W. State-of-health estimation of lithium-ion batteries using a novel dual-stage attention mechanism based recurrent neural network. *J. Energy Storage* **72**, 109297 (2023). https://doi.org/10.1016/j.est.2023.109297
[30] Oza, M., Lamba, R. & Kumar, R. Attention-based mechanism for state-of-health estimation of Li-ion batteries. *Proc. 2024 IEEE 4th Int. Conf. Sustainable Energy and Future Electric Transportation (SEFET)*, 1–6 (2024). https://doi.org/10.1109/SEFET61574.2024.10718141
[31] Wang, Q. & Yang, S. State-of-health estimation for lithium-ion batteries based on BiLSTM-WGAN. *Proc. 2023 9th Int. Conf. Computer and Communications (ICCC)*, 1541–1547 (2023). https://doi.org/10.1109/ICCC59590.2023.10507596
[32] Eddin, E. E. E., Massaoudi, M. & Ghrayeb, A. Accurate state of health estimation of lithium-ion batteries: An efficient ResNet-LSTM approach. *Proc. 2024 4th Int. Conf. Smart Grid and Renewable Energy (SGRE)*, 1–6 (2024). https://doi.org/10.1109/SGRE59715.2024.10428975
[33] Feng, S., Song, M., Lin, Y., Yao, W., and Xie, J. CNN–LSTM-based state of health estimation for Li-ion batteries under multiple working conditions. *Energy Technol.* **12**, 202301039 (2023). https://doi.org/10.1002/ente.202301039
[34] Sun, S., Sun, J., Wang, Z., Zhou, Z., and Cai, W. Prediction of battery SOH by CNN-BiLSTM network fused with attention mechanism. *Energies* **15**, 4428 (2022). https://doi.org/10.3390/en15124428
[35] Alhussein, M., Aurangzeb, K. & Haider, S. I. Hybrid CNN-LSTM model for short-term individual household load forecasting. *IEEE Access* **8**, 180544–180557 (2020). https://doi.org/10.1109/ACCESS.2020.3028281
[36] Gao, D., Liu, X., Zhu, Z., and Yang, Q. A hybrid CNN-BiLSTM approach for remaining useful life prediction of EVs lithium-ion battery. *Meas. Control* **56**, 3–13 (2023). https://doi.org/10.1177/00202940221103622
[37] Zhao, G. *et al.* Capacity prediction and remaining useful life diagnosis of lithium-ion batteries using CNN-LSTM hybrid neural network. *Proc. 2021 China Automation Congress (CAC)*, 2179–2183 (2021). https://doi.org/10.1109/CAC53003.2021.9727496
[38] Cai, Z., Fan, Q., Feris, R. S., Vasconcelos, N. A unified multi-scale deep convolutional neural network for fast object detection. *ECCV 2016*, 354–370 (2016).
[39] Cui, Z., Chen, W. & Chen, Y. Multi-scale convolutional neural networks for time series classification. arXiv:1603.06995 (2016).

ARTICLE IN PRESS

ARTICLE IN PRESS

[40] Nah, S., Kim, T. H. & Lee, K. M. Deep multi-scale convolutional neural network for dynamic scene deblurring. *Proc. CVPR*, 3883–3891 (2017).
[41] Cooijmans, T., Ballas, N., Laurent, C., Gülçehre, Ç., and Courville, A. Recurrent batch normalization. arXiv:1603.09025 (2016).
[42] Santurkar, S., Tsipras, D., Ilyas, A., Madry, A. How does batch normalization help optimization? *Adv. Neural Inf. Process. Syst.* **31** (2018).
[43] Bjorck, N., Gomes, C. P., Selman, B., Weinberger, K. Q. Understanding batch normalization. *Adv. Neural Inf. Process. Syst.* **31** (2018).
[44] Larsson, G., Maire, M. & Shakhnarovich, G. FractalNet: Ultra-deep neural networks without residuals. arXiv:1605.07648 (2016).
[45] Zhang, J., Zhu, Y., Li, W., Fu, W., Cao, L. DRNet: A deep neural network with multi-layer residual blocks improves image denoising. *IEEE Access* **9**, 79936–79946 (2021).
[46] Wu, Z. *et al.* Blockdrop: Dynamic inference paths in residual networks. *Proc. CVPR*, 8817–8826 (2018).
[47] Xie, S., Girshick, R., Dollár, P., Tu, Z., He, K. Aggregated residual transformations for deep neural networks. *Proc. CVPR*, 1492–1500 (2017).
[48] Cheng, J., Dong, L. & Lapata, M. Long short-term memory-networks for machine reading. arXiv:1601.06733 (2016).
[49] Staudemeyer, R. C. & Morris, E. R. Understanding LSTM: A tutorial into long short-term memory recurrent neural networks. arXiv:1909.09586 (2019).
[50] Le, X. H., Ho, H. V., Lee, G., Jung, S. Application of long short-term memory (LSTM) neural network for flood forecasting. *Water* **11**, 1387 (2019).
[51] Kratzert, F., Klotz, D., Brenner, C., Schulz, K., Herrnegger, M. Rainfall–runoff modelling using LSTM networks. *Hydrol. Earth Syst. Sci.* **22**, 6005–6022 (2018).
[53] Laghrissi, F., Douzi, S., Douzi, K., Hssina, B. Intrusion detection systems using long short-term memory (LSTM). *J. Big Data* **8**, 65 (2021).
[53] Javid, A. M., Douzi, S., Douzi, K., Hssina, B. A ReLU dense layer to improve the performance of neural networks. *Proc. ICASSP 2021*, 2810–2814 (2021).
[54] Jeczmionek, E. & Kowalski, P. A. Flattening layer pruning in convolutional neural networks. *Symmetry* **13**, 1147 (2021).
[55] Chieng, H. H., Wahid, N., Ong, P., Perla, S. R. K. Flatten-T Swish: A thresholded ReLU-Swish-like activation function for deep learning. arXiv:1812.06247 (2018).
[56] Eger, S., Youssef, P. & Gurevych, I. Is it time to swish? Comparing deep learning activation functions across NLP tasks. arXiv:1901.02671 (2019).
[57] Ide, H. & Kurita, T. Improvement of learning for CNN with ReLU activation by sparse regularization. *Proc. IJCNN 2017*, 2684–2691 (2017).
[58] Eckle, K. & Schmidt-Hieber, J. A comparison of deep networks with ReLU activation function and linear spline-type methods. *Neural Netw.* **110**, 232–242 (2019).
[59] Schmidt-Hieber, J. Nonparametric regression using deep neural networks with ReLU activation function. (2020).

ARTICLE IN PRESS

[60] Sharma, S., Sharma, S. & Athaiya, A. Activation functions in neural networks. *Towards Data Science* **6**, 310–316 (2017).

ARTICLE IN PRESS