ORIGINAL ARTICLE

Application of CIGAR for collective learning between CNN–BiLSTM models in lithium-ion battery state of health prediction

Sylwia Olbrych · Zi Xuan Tung · Sehriban Celik · Hans Aoyang Zhou · Anas Abdelrazeq · Dirk Uwe Sauer · Robert H. Schmitt

Received: 27 January 2025 / Accepted: 10 April 2026

Abstract
Battery management systems (BMSs) are essential for accessing and managing battery performance information, with state of health (SOH) estimation providing insights into the battery's life expectancy. Electrochemical impedance spectroscopy (EIS) is a non-destructive method for SOH assessment. However, collecting EIS data across diverse operating conditions and battery types is both time-intensive and costly, presenting challenges related to data distribution and heterogeneity. This work investigates a lightweight gradient-based fusion strategy to enable collective learning across independently trained models without sharing raw data. Specifically, the collective inference via gradient aggregation (CIGAR) algorithm is applied to multiple convolutional neural network–bidirectional long short-term memory models trained on disjoint EIS datasets. The approach is evaluated on a real-world SOH prediction task, demonstrating that gradient-based collective learning can facilitate knowledge exchange among models under compatible conditions. The results highlight both the potential and the limitations of CIGAR in heterogeneous battery scenarios, indicating that further optimisation and validation on larger and more diverse datasets are required to improve robustness and generalisation.

Keywords Machine learning · Collective intelligence · State of health · Electrochemical impedance spectroscopy · Lithium-ion battery

1 Introduction

Due to their high energy density and fast power delivery, lithium-ion batteries (LIBs) play a crucial role in the energy storage sector, making them the technology of choice for applications such as electric vehicles (EVs), portable electronics, and grid-scale energy storage [16, 28]. Recognising their importance in the battery industry, the electronics sector has increasingly focused on researching and improving the reliability of LIBs. One of the critical parameters in assessing the capacity, performance, and lifetime of LIBs is their state of health (SOH) [13]. Reliable SOH predictions can be integrated into a battery management system (BMS) to prevent malfunctions and performance deterioration by scheduling maintenance ahead of time or sending warnings before defects

Published online: 27 May 2026

reach critical levels [18]. Electrochemical impedance spectroscopy (EIS) has proven to be an effective real-time, non-invasive technique that provides detailed insights into a battery's electrochemical processes by exposing changes in the battery impedance over a range of probing voltage or current frequencies [14]. Through examining EIS data, researchers can extract features indicative of battery ageing and use them to estimate the SOH [17]. Therefore, building machine learning (ML) models to correlate changes in EIS spectra with battery degradation has significant potential for predicting battery health [11] and incorporation into an advanced BMS.

The current literature reveals the gap in investigating collective intelligence techniques that can enable collaboration to extract the full potential of EIS measurements. Traditional SOH prediction models are often limited by the variability in measurement data, resulting in restricted generalizability and performance in real-world applications [3]. The gradient-based fusion approach applied in this work allows individual entities, such as stakeholders, testing centres and laboratories, to exchange learned patterns from training data collected under similar measurement conditions without directly sharing the underlying model parameters. This study focuses on enabling collaboration under practical constraints, such as data privacy and heterogeneous measurement conditions. Following this approach can reduce the reliance on extensive individual testing campaigns while providing a principled mechanism for collective inference across distributed data sources. The key contributions of this work are:

- The application of the collective inference via gradient aggregation (CIGAR) algorithm as a mechanism for enabling collective learning among independently trained models without exchanging raw data or model parameters.
- An empirical analysis of the applicability and limitations of CIGAR-based fusion when combining models trained on heterogeneous EIS datasets.
- A statistically validated evaluation of the fusion process, including an analysis of convergence behaviour and the influence of fusion iterations under controlled experimental conditions.

The remainder of this paper is structured as follows. In Section 2, research on data-driven methods for EIS-based SOH prediction is summarised. Section 3 introduces the proposed methodology, including details on data acquisition, data preparation, collective learning algorithm, and the evaluation process. Section 4 presents the results obtained from applying the CIGAR approach, followed by the discussion in Section 5. Finally, Section 6 concludes the paper and outlines potential future research directions.

2 Related work

This section reviews representative studies relevant to EIS-based SOH estimation and collective learning under heterogeneous settings, rather than providing an exhaustive survey of all deep learning architectures used for SOH prediction.

Estimating the SOH for LIBs has become a focal point in electromobility and stationary storage applications. As the demand for batteries continues to grow, recent research increasingly focuses on EIS and ML techniques to tackle the challenges posed by data diversity and the complexity of battery processes. To build on this progress, several studies have investigated the integration of an equivalent circuit model (ECM) of batteries with probabilistic and ML methods and presented various approaches to SOH prediction under different operating conditions.

Studies, such as the one by Zhang et al. [31], have developed probabilistic models that integrate ECM parameters for SOH estimation, explicitly considering the effects of temperature and state of charge (SOC) variations. Li et al. [13] further expanded on this by incorporating ECM parameters into ML models, specifically using Gaussian process regression (GPR) to enhance the model's sensitivity to temperature changes. Additionally, Rastegarpanah et al. [24] applied neural networks for SOH estimation of LiBs, utilising parameters extracted from impedance data modelled through a modified Randles ECM.

The ML techniques have also been widely applied for SOH estimation by directly utilising EIS data, eliminating the need for traditional ECMs. For instance, Rastegarpanah et al. [25], applied a convolutional neural network (CNN) designed to operate even when trained with limited data. Faraji Niri et al. [4] adopted a GPR model similar to Li et al. [13], but without incorporating an ECM and instead focused on the influence of temperature and SOC values on prediction accuracy. To analyse the performance of their GPR model for SOH prediction, they conducted two case studies: one using all the data points from the EIS spectra along with SOC and temperature as model inputs, and the other using specific features extracted from the EIS plots as inputs. In another approach, Obregon et al. [22] developed a hybrid architecture combining a convolutional autoencoder (CAE) with a deep neural network (DNN) to predict SOH based on EIS measurements. Guo et al. [8] introduced a unique method of transforming EIS data into images using the Gramian angular field technique, followed by feature extraction with a VGG16 neural network. By integrating automatic feature extraction and SOH prediction into a single neural network, this study minimised the need for manual feature engineering. Li et al. [14] combined CNNs with bidirectional long short-term memory (BiLSTM) networks, optimised by particle swarm optimisation. Similarly, Zhang et al. [29] investigated the convolutional neural network bidirectional long short-term memory (CNN–BiLSTM) architecture for EIS-based SOH estimation.

These studies, summarised in Table 1, illustrate the trend toward combining EIS data with advanced ML and data-driven techniques for predicting SOH. The main focus is on models adapting to varying temperatures and limited data, which is crucial for advancing BMS. In this study, a CNN–BiLSTM architecture is applied directly to EIS and capacity data, eliminating the need for ECM integration.

To bring theoretical advancements into praxis, this study examines the applicability of algorithmic innovation to facilitate collaborative intelligence in real-world scenarios. Some prior works such as [7, 10, 30] have explored solutions which rely on the assumption of complete transparency of local models, requiring fully accessible model architectures and openly shared local training data. On the other hand, Hoang et al. [9] introduced a solution which addresses the challenge of collective model fusion for experts with heterogeneous black-box architectures. The authors also identified the need for a decentralised learning paradigm to perform fusion without direct access to black-box training data and architectures and avoid centralised bottlenecks. According to the survey on deep model fusion by Li et al. [15], this approach falls under the category of ensemble learning by combining diverse models' outputs to improve the final prediction accuracy and robustness. To enable decentralised collective learning without sharing training data or model parameters, this study implements the CIGAR approach proposed by Hoang et al. [9].

Table 1 Overview of various approaches for predicting SOH using EIS data

Study	Machine learning	Diverse temp.	With ECM	Diverse design	Collective learning
Zhang et al. [29]	✓	✓	-	-	-
Rastegarpanah et al. [24]	✓	-	✓	-	-
Rastegarpanah et al. [25]	✓	-	-	-	-
Zhang et al. [31]	-	✓	✓	-	-
Li et al. [14]	✓	✓	✓	-	-
Faraji-Niri et al. [4]	✓	✓	-	-	-
Guo et al. [8]	✓	✓	-	-	-
Obregon et al. [22]	✓	✓	-	-	-
Li et al. [13]	✓	✓	✓	-	-
Liu et al. [19]	✓	✓	-	-	-
Our	✓	-	-	✓	✓

3 Methods

From related work, it is apparent that effectively addressing the challenges of data-driven collaborative SOH prediction requires advanced techniques capable of handling diverse, decentralised datasets. Accordingly, this section begins with data acquisition, providing a detailed description of the characteristics of the utilised datasets. The methodology for data processing and developing the CNN–BiLSTM models, designed for SOH prediction of LIBs, is then outlined. This is followed by an introduction to the CIGAR algorithm and a discussion of the evaluation metrics used to assess the CIGAR performance.

3.1 Data acquisition

Dataset A was obtained from the work of Zhang et al. [32], where the data are publicly available. This dataset includes detailed capacity characteristics and EIS measurements for 45 mAh Eunicell LR2032 lithium-ion coin cells. A total of eight cells tested at 25 °C (labelled A01 to A08) were used in this study. The cells were cycled using a 1 C constant-current (CC), constant-voltage (CC-CV) charging method and a 2 C CC discharge method. The EIS measurements were conducted over a frequency range of 0.02 Hz to 20 kHz, with an excitation current of 5 mA. The discharge cutoff voltage was set at 3 V. This data covers situations essential to comprehend the behaviour of the cells, such as relaxed and non-relaxed states and various times during the cycles of charging and discharging. The SOC of 100% within a 15-minute rest period was selected in this study as it provides the most optimal condition for measurements [12]. This choice guarantees that the data accurately depicts the performance of the cells under normal usage circumstances and permits the observation of any variations in impedance, especially in the low-frequency range, which indicates a change in cell dynamics.

The EIS measurements for LIBs in Dataset A are organised into distinct stages, each representing a specific phase in the battery's charging or discharging cycle. In Dataset A, a DC offset is applied in stages II, III, VI, and VII, while resting periods are incorporated in stages I, II, V, VI, and IX. Additionally, the availability of EIS measurements in each stage was evaluated. As seen in Table 2, stages VI, VII and VIII are not complete, and including them in model training can disrupt the capacity of the model to learn patterns across stages, potentially resulting in worse prediction accuracy or over-fitting to the available data [5]. Therefore, these stages are excluded from the final dataset.

Subsequently, Nyquist plots for stages I, II, III, IV, V, and IX were analysed to identify the most representative data. The plots were generated using data comprising the real part ($\mathrm{Re}(\underline{Z})$), imaginary part ($\mathrm{Im}(\underline{Z})$), frequency, along with their corresponding cycle numbers (see Fig. 1). Stages II, III, and IV exhibited low-frequency anomalies due to the absence of a resting period and the presence of DC during measurements, highlighting the sensitivity of EIS to test conditions. In contrast, stages I, V, and IX, which included a 15-minute rest period, showed more stable behaviour. Since stages I and IX are at 0% SOC and EVs typically avoid reaching 0% SOC to reduce battery ageing [26], stage V was chosen for model training.

Table 2 Presence of EIS data at different stages across cells in Dataset A

Cell	Stage								
	I	II	III	IV	V	VI	VII	VIII	IX
A01	✓	✓	✓	✓	✓	✓			✓
A02	✓	✓	✓	✓	✓	✓			✓
A03	✓	✓	✓	✓	✓	✓			✓
A04	✓	✓	✓	✓	✓		✓	✓	✓
A05	✓	✓	✓	✓	✓	✓			✓
A06	✓	✓	✓	✓	✓	✓	✓	✓	✓
A07	✓	✓	✓	✓	✓	✓	✓	✓	✓
A08	✓	✓	✓	✓	✓	✓	✓	✓	✓

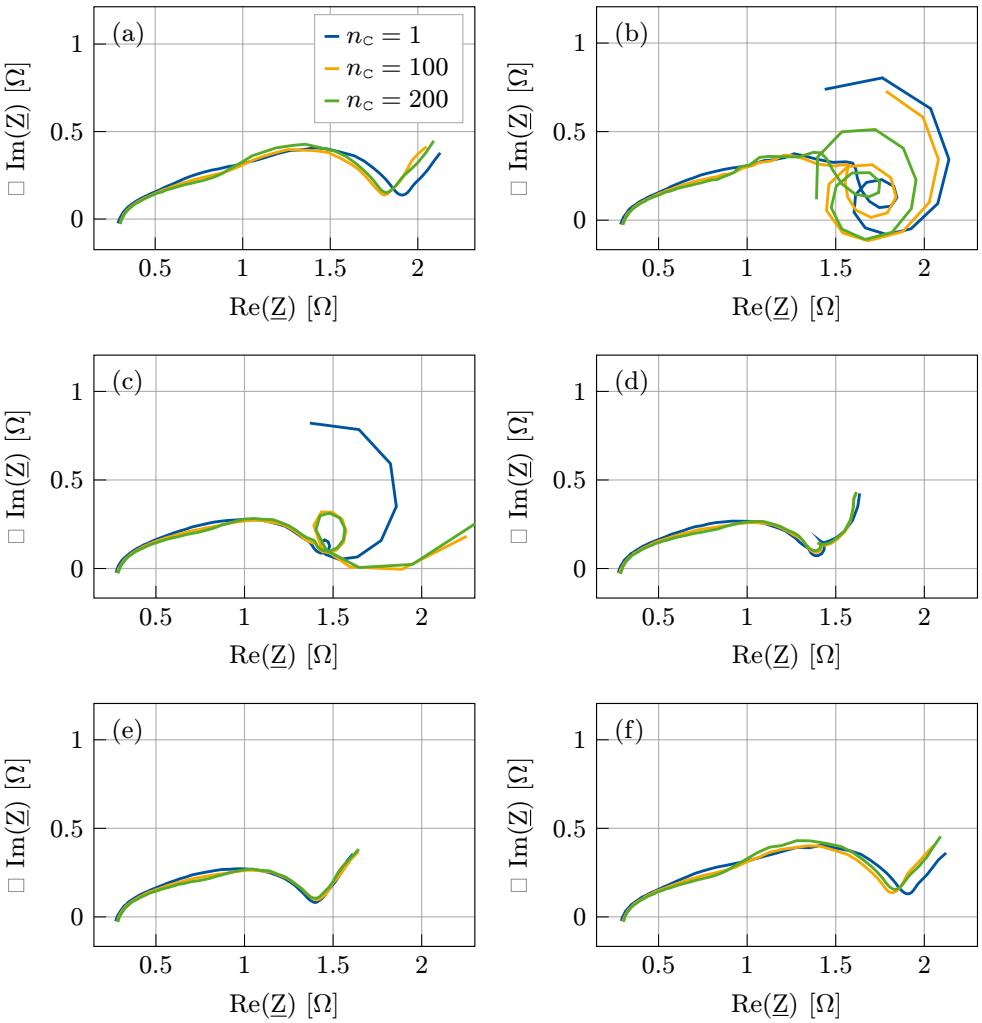

Fig. 1 Nyquist plots of cell A02 in Dataset A at different cycles across various stages (**a**) stage I, (**b**) stage II, (**c**) stage III, (**d**) stage IV, (**e**) stage V and (**f**) stage IX

The values labelled *ox/red* in Dataset A were utilised to distinguish between the charge and discharge phases of the capacity. In this context, reduction and oxidation reactions represent the charging and discharging states of the cells, where *ox/red* = 1 corresponds to charging and *ox/red* = 0 to discharge. The discharge data was then analysed to determine the maximum capacity value in the discharge cycle. Subsequently, the initial capacity was extracted for SOH calculation using the ratio of the current maximum capacity of the battery (C_m) to its initial capacity (C_{init}) at its new state [13]:

$$SOH = \frac{C_m}{C_{init}} \times 100. \tag{1}$$

In Fig. 2, the normalised SOH is plotted against the cycle number to evaluate the suitability of each cell in stage V for processing. It was observed that, in contrast to the other cells, the A04 and A08 cells have a significantly shorter cycle number, reaching only about 40 cycles. To maintain consistency in the data and ensure that all cells are within the same operational range, the data from cells A04 and A08 were excluded from the model training.

Dataset B consists of three lithium-ion pouch cells identified as B01, B02 and B03, all measured at a controlled temperature of 25 °C. The nominal voltage of these cells is 3.6 V, with a rated capacity of 14.0 Ah, a maximum voltage of 4.2 V, and a discharge cutoff voltage of 2.9 V. EIS measurements were conducted over a frequency range of 0.01 Hz to 6 kHz, with the charge/discharge cycle using a 1 C CC-CV charging protocol and a 1 C CC discharging protocol. For this dataset, EIS tests were conducted at 20 %, 50 % and 80 % SOC levels. Given that

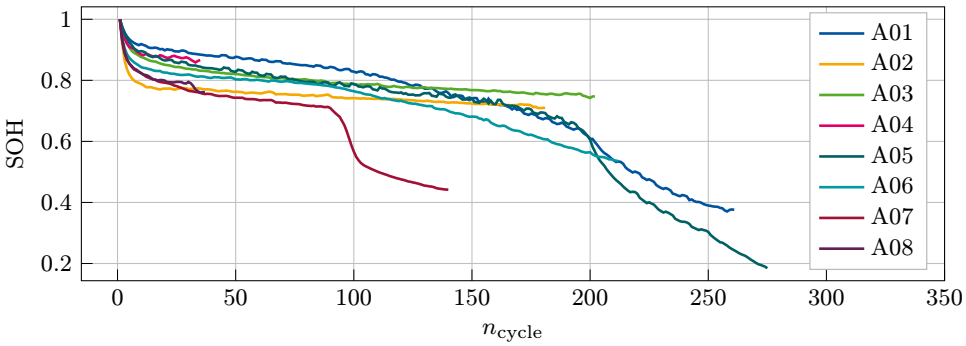

Fig. 2 SOH degradation profiles of cells at stage V in Dataset A as a function of cycle number

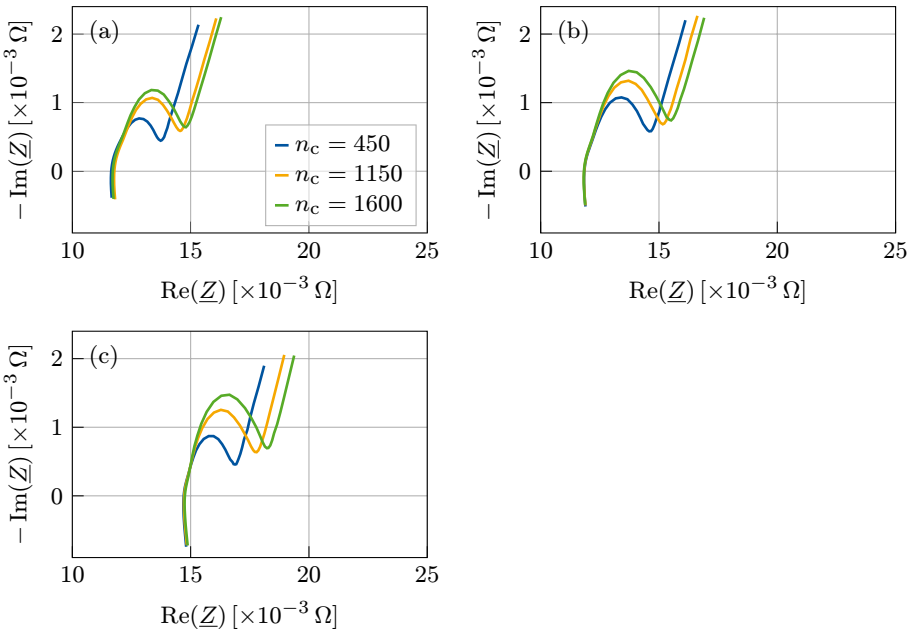

Fig. 3 Nyquist plots of cells (**a**) B01, (**b**) B02 and (**c**) B03 in Dataset B at different cycle numbers

Dataset A uses the SOC of 100 %, the SOC of 80 % from Dataset B was selected as the most comparable to the SOC conditions of cells at stage V from Dataset A. To construct the Nyquist plots for further analysis, a similar procedure as for Dataset A was carried out (see Fig. 3). The frequency points above 800 Hz were excluded from the plots and dataset, as the Nyquist line exhibited a downward trend at high frequency and crossed below 0 on the y-axis. The section where such an inductive behaviour is observed is often considered irrelevant for estimating battery capacity as it may be an artefact of the measurement system [21], or may originate from the geometry of the cell and its windings [27]. After this exclusion, cells in Dataset B have 41 frequency points, resulting in 41 values for $\text{Re}(\underline{Z})$ and $\text{Im}(\underline{Z})$ for each cycle.

Maximum capacity values in Dataset B from each discharge cycle, along with the corresponding cycle numbers, were extracted to calculate SOH using Eq. 1. The SOH degradation of the pouch cells in Dataset B was then analysed, as shown in Fig. 4. It is evident that all cells underwent approximately the same number of cycles without abnormal SOH values, and as a result, no data points were excluded from the analysis.

Dataset C contains custom-manufactured 5 Ah NMC811/graphite pouch cells examined by Mohtat et al. [20] at the University of Michigan Battery Lab. A total of 31 cells were manufactured in a single batch and subjected to a series of controlled ageing experiments. Although the complete dataset comprises 31 cells, this study focuses on the 21 cells that were cycled under varying temperature, C-rate, and depth of discharge conditions. These cells were aged at −5 °C, 25 °C, and 45 °C, using CC-CV charge and CC and drive-cycle discharge protocols to a

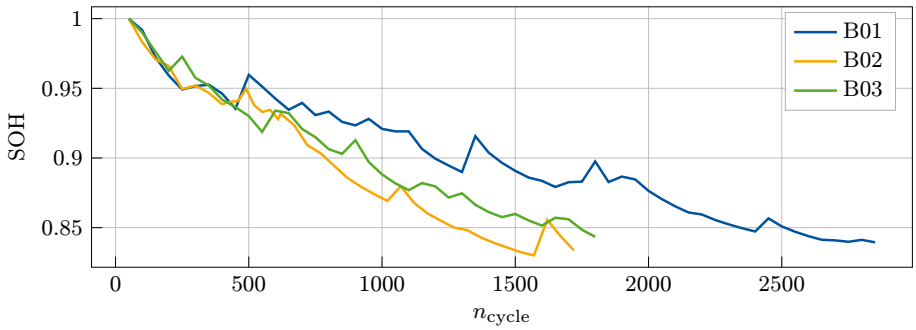

Fig. 4 SOH degradation profiles of cells in Dataset B as a function of cycle number

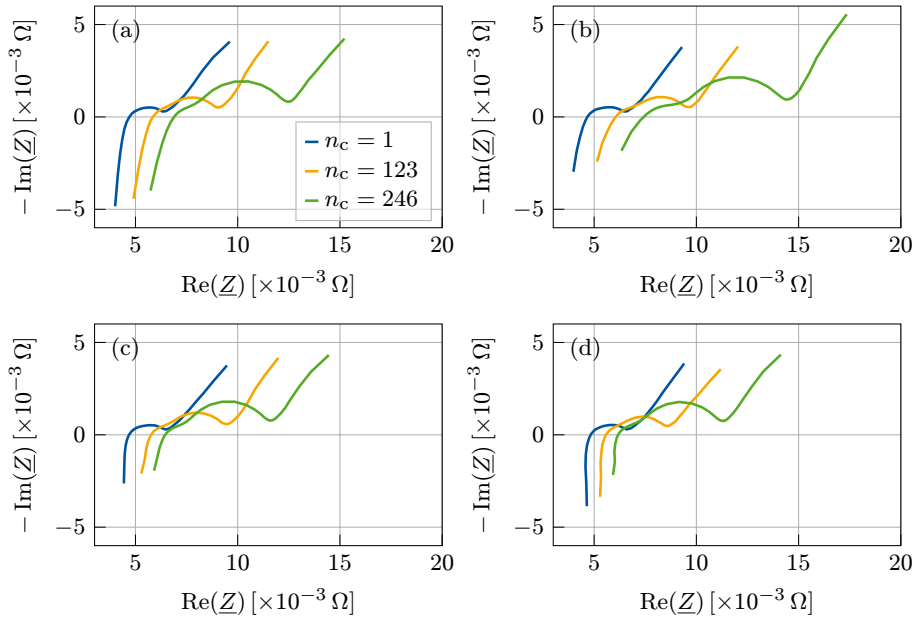

Fig. 5 Nyquist plots of cells (**a**) C04, (**b**) C07, (**c**) C10, and (**d**) C19 in Dataset C at different cycle numbers

cut-off voltage of 3 V. Several operating conditions were applied, including charge and discharge rates of 0.2C, 1.5C, and 2 C, as well as a representative drive-cycle discharge case. EIS measurements were performed as part of reference performance tests at 25 °C, regardless of the cyclic ageing temperature. For cells aged at −5 °C or 45 °C, the cells were returned to room temperature and allowed to rest for at least 3 hours to ensure thermal equilibrium prior to impedance measurements. The EIS spectra were recorded over a frequency range of 10 mHz to 10 kHz at 10% SOC intervals, following a standardised protocol comprising discharge, rest periods, and hybrid pulse power characterisation at each SOC level. For modelling purposes, the highest available SOC of 90% was selected.

In this study, the preprocessed and curated version of this dataset provided by Blömeke et al. [2] was used, which retains the original data structure while providing analysis-ready EIS files for SOH estimation. Representative Nyquist plots are presented in Fig. 5, indicating the suitability of this dataset for modelling.

The SOH profiles of the pouch cells in Dataset C were analysed as a function of cycle number, as shown in Fig. 6. During preprocessing, outliers that could adversely affect the modelling were excluded. No additional preprocessing steps were required.

Table 3 compares the features of the three datasets, highlighting both their similarities and differences. The key distinction lies in the type of cells and their rated capacities. Despite these differences, the datasets share several complementary characteristics, making them well-suited for investigating collective learning. The datasets were generated using the same charging procedure, with identical maximum voltage and a comparable frequency

 https://doi.org/10.1007/s00521-026-12144-6

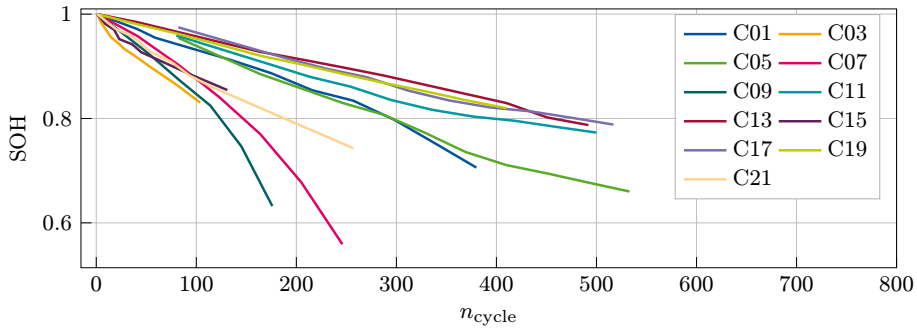

Fig. 6 SOH degradation profiles of cells in Dataset C as a function of cycle number

Table 3 Summary of datasets used in this study and corresponding cell specifications

Feature	Dataset A	Dataset B	Dataset C
Data source	Zhang et al. 2020 [32]	In-house measurements	Mohtat et al. 2021 [20]
Cell type	Li-ion coin cells	Li-ion pouch cells	Li-ion pouch cells
Cell chemistry	$LiCoO_2$ Graphite	$LiNi_{0.6}Mn_{0.2}Co_{0.2}O_2$ Graphite	$LiNi_{0.8}Mn_{0.1}Co_{0.1}O_2$ Graphite
Cell number	8	3	21
Temperature (°C)	25	25	25
Rated capacity (Ah)	0.045	14.0	5.0
Nominal voltage (V)	-	3.6	-
Max voltage (V)	4.2	4.2	4.2
Discharge cutoff voltage (V)	3.0	2.9	3.0
Frequency (Hz)	0.02–20 000	0.01–6 000	0.01–10 000
Charge	1 C CC-CV	1 C CC-CV	CC-CV at 0.2C/1.5C/2C
Discharge	2 C CC	1 C CC	CC at 0.2C/1.5C/2 C; drive-cycle discharge
Operating conditions	Controlled charge-discharge cycle; relaxed and non-relaxed states	Controlled charge-discharge cycle; relaxed states	Controlled charge-discharge cycle; relaxed states

Table 4 Flattened EIS data with frequency values as columns

	f_1	f_2	...	f_n
$\mathrm{Re}(\underline{Z})$	−0.022220	0.022436	...	0.47839
$-\mathrm{Im}(\underline{Z})$	0.32554	0.41503	...	1.8269

range. Since datasets feature graphite anodes and lithium-based cathodes, the underlying chemical processes during charging and discharging are largely similar. Furthermore, all measurements were conducted at 25 °C.

3.2 Data processing

To properly utilise EIS data as input for a neural network, it must be transformed into a suitable format. This transformation involves structuring the data so that each row represents the real part $\mathrm{Re}(\underline{Z})$ and imaginary part $\mathrm{Im}(\underline{Z})$ of impedance values, with frequency points displayed in separate columns (see Table 4). The steps applied in data processing are as follows:

1. **Data import:** Load the capacity and EIS data.
2. **Data selection:** Select $\mathrm{Re}(\underline{Z})$, $\mathrm{Im}(\underline{Z})$, and the target variable *capacity*. Append the extracted $\mathrm{Re}(\underline{Z})$ and $\mathrm{Im}(\underline{Z})$ values to the created *eis* array.

3. **Capacity data preparation:** Isolate the initial discharge capacity from the dataset and apply Eq. 1 to calculate the SOH. Append the resulting SOH values to the created *soh* list. Reshape the *soh* list into a 2D array and scale it using `MinMaxScaler`:

$$X_{\text{scaled}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}}. \tag{2}$$

4. **EIS data preparation:** Extract the impedance values, flatten, and reshape $\mathrm{Re}(\underline{\mathrm{Z}})$ from the *eis* array into a 2D format. Scale values using `MinMaxScaler` by computing each feature's minimum and maximum values. After scaling, return the data to its original dimensions by reshaping and concatenating the scaled $\mathrm{Re}(\underline{\mathrm{Z}})$ to a dictionary. Repeat this process for $\mathrm{Im}(\underline{\mathrm{Z}})$ to ensure that both parts are prepared consistently.
5. **Return:** Finally, return the transformed *DataFrame* and the *Scaler*. This step ensures the data is ready to be input into the neural network, facilitating effective training and validation.

The resulting *DataFrame* is a structured dictionary of the scaled features and targets for each battery cell. Key steps include importing and selecting data, calculating SOH using initial capacity, transforming data into a 2D format, scaling values with `MinMaxScaler`, and organising the processed data into a *DataFrame* for neural network input.

3.3 Model development

As a next step, three independent CNN–BiLSTM models were developed, whose architecture can be found in Table 5. The CNN–BiLSTM architecture was selected based on literature research, as it demonstrates promising performance for predicting SOH from EIS data [19]. It is important to note that *n* represents the number of cycles, which varies across different cells. The model first applies a 1D convolutional layer (`Conv1d`) for feature extraction. Since datasets have a varying number of features, an additional convolutional layer `Conv1D (adjust)` is applied. This adjustment ensures that datasets have the same feature length, required for gradient fusion via CIGAR.

Each dataset was split into training and test sets to facilitate K-fold cross-validation and model performance testing. For Dataset A, coin cells A01, A02, A03, A06 and A07 were selected to form the training set, while coin cell A05 was allocated to the test set. For Dataset B, pouch cells B01 and B02 were used for training, while B03

Table 5 Implemented CNN–BiLSTM model architecture

Layer type	Output shape	Param#	Hyperparameters
Permutation	[n, 2, 60]	–	–
Conv1d	[n, 2, 60]	6	kernel_size = 1
Conv1d (adjust)	[n, 2, 41]	82	kernel_size = 20, stride = 1
Permutation	[n, 41, 2]	–	–
LSTM	[n, 41, 256]	1,716,224	hidden_size = 128, num_layers = 5 bidirectional = True
Dropout	[n, 41, 256]	–	p = 0.3
Linear	[n, 1]	257	–
Total params:		1,716,569	
Trainable params:		1,716,569	
Non-trainable params:		0	

was assigned as the test set. For Dataset C, pouch cells from C01 to C20 were used for training and following C21 was left for the test set. The K-fold cross-validation was employed due to the relatively small size of the dataset, which helps to prevent overfitting [23]. For this procedure, the training set was divided into five distinct folds using random shuffling. For each iteration, four folds (80%) were used for training and one fold (20%) for validation, resulting in five iterations in total. The number of iterations ensured that each fold served once as the validation set. After completion of model training, the test data, which was initially set aside, was input into each of the five trained models. The mean of the outputs from these models was then used to calculate the final prediction.

3.4 Collective learning

The CNN–BiLSTM models developed in this work function as specialised experts within a collective learning framework, with each model operating independently. The predictions y_i, represented by p_i, the likelihood of observing class i given input x_i as defined in Eq. 3, are fused using the CIGAR gradient-based algorithm. This fusion process progressively refines each expert's prediction accuracy over multiple iterations [9].

$$y_i = p_i(x_i|l_i, \boldsymbol{\omega}_i) \tag{3}$$

This study demonstrates the learning approach on datasets, where one model receives learned gradients from other participants and *vice versa*. The learning experts in the framework are designed to remain consistently available for querying. Figure 7 illustrates the author's understanding and implementation of the CIGAR algorithm.

The CIGAR algorithm is used for lightweight gradient fusion, where each predictor $p_i(\ell_i\ \boldsymbol{\omega}_i)$ denotes the local expert parameterised by the non-linear function ℓ_i and its characterising parameters $\boldsymbol{\omega}_i$ associated with i^{th} data point. The local log-likelihood prediction gradient can then be represented as $\nabla \log p_i(\ell_i; \boldsymbol{\omega}_i)$. Assuming access to local gradients of m experts, one can compute the global prediction gradient via:

$$\nabla \log p = \sum_{i=1}^{m} \nabla \log p_i\left(\ell_i; \boldsymbol{\omega}_i\right) \tag{4}$$

The framework determines $\log p_i$ based on its mean μ_i and variance σ_i^2, representing a measure of the likelihood that input x_i outputs the observed value y_i using:

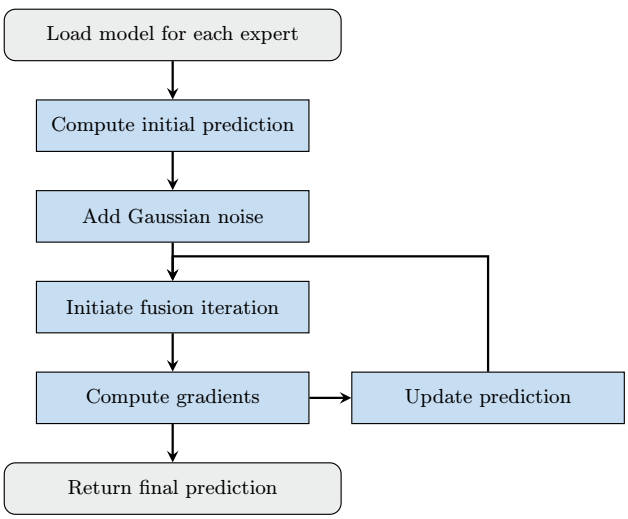

Fig. 7 Flowchart of our implementation of CIGAR-based fusion

$$\log p_i(y) = \log\left(\frac{1}{\sqrt{2\pi}\cdot\sqrt{\sigma_i^2}}\right) - \frac{(y_i-\mu_i)^2}{2\cdot\sigma_i^2} \tag{5}$$

When experts cannot disclose their models' architectures, the noise perturbation can be added to estimate the gradient numerically. This involves adding Gaussian noise to the predictions and then calculating the gradient by observing how the predictions change in response to this perturbation. The presented approach enables the model to capture subtle relationships between predictions and adapt accordingly. In CIGAR, the primary role of noise is to enhance the fusion of predictions from different models by adjusting the gradients of the likelihood. Since directly computing $\nabla_y \log p_i\left(y \mid \ell_i; \boldsymbol{\omega}_i\right)$ may be challenging in such settings, a random gradient prediction method is employed:

$$\begin{aligned}\nabla_y \log p_i\left(y\right) &\simeq \mathbb{E}_z\left[\frac{z}{\lambda_y}\log\left(\frac{p_i\left(y+\lambda_y z\right)}{p_i\left(y\right)}\right)\right] \\ &= \mathbb{E}_z\left[\nabla_y^{(z)}\log p_i\left(y\right)\right]\end{aligned} \tag{6}$$

The term involves Gaussian noise z, sampled from a standard normal distribution $z \sim \mathcal{N}(0,1)$. This noise is introduced to perturb the original prediction y. The perturbation $\lambda_y z$ means that the original prediction y is shifted by a small amount, proportional to the noise z and the scaling factor λ_y, which ensures that the perturbation is small enough not to alter the prediction drastically. Equation 7 represents an approximation of the log-likelihood gradient for a single noise realisation, denoted by z.

$$\nabla_y^{(z)}\log p_i(y) \triangleq \frac{z}{\lambda_y}\log\left(\frac{p_i\left(y+\lambda_y z\right)}{p_i(y)}\right) \tag{7}$$

To reach a consensus on predictions, each expert's local prediction estimate $y_i = y_i^{(t)}$ is updated using the global prediction gradient (Equation 4) and gradient ascent, with an adequately small learning rate $\eta > 0$.

$$y_i^{(t+1)} = y_i^{(t)} + \eta\nabla_y \log p\left(y^{(t)} \mid \mathrm{x}\right) \tag{8}$$

To ensure a controlled experimental setting, all fusion computations are performed centrally on a single machine. Let b denote the number of participating experts and T the number of fusion iterations. In each iteration, the CIGAR-based fusion algorithm queries all experts to estimate gradient contributions via numerical perturbations and updates the fused prediction vector. Treating each expert as a black box, the dominant computational cost per iteration is proportional to the number of forward predictions required for gradient estimation. The overall time complexity is therefore $\mathcal{O}(T \times b \times n \times n_z \times C_{\text{pred}})$, where n is the number of test samples, n_z the number of perturbation samples used in gradient estimation, and C_{pred} the cost of a single model inference. The memory complexity scales as $\mathcal{O}(n)$, since only prediction vectors are stored, while model parameters remain fixed.

3.5 Evaluation

The evaluation protocol is designed to assess the performance of the baseline models, the uniform ensemble, and the CIGAR-based fusion approach. The independently trained expert models were evaluated using a K-fold cross-validation scheme to establish baseline generalisation performance. The CIGAR fusion mechanism was evaluated through a grid search over the number of fusion iterations. Each fusion configuration was explored over 40 fusion iterations, with each iteration repeated 10 times, resulting in a total of 400 runs per fusion setup. Performance metrics were averaged across repetitions for each fusion iteration number to obtain average performance estimates.

Root Mean Squared Error (RMSE) and the coefficient of determination (R^2) were used as evaluation metrics. RMSE, a widely used metric for regression tasks, computes the square root of the average squared differences between predicted and true values [6]:

$$\mathrm{RMSE}(\mathrm{X}, h) = \sqrt{\frac{1}{m}\sum_{i=1}^{m}\left(h\left(\mathrm{x}^{(i)}\right) - y^{(i)}\right)^2} \tag{9}$$

The coefficient of determination (R^2) indicates how well model predictions explain the variance in the ground-truth data, with higher values corresponding to stronger agreement between predicted and actual values [1, 4].

4 Results

This section presents the results of the applied CIGAR-based fusion approach. The analysis is structured to first examine the behaviour of the fusion process across iterations, and then compare the final fusion outcomes against baseline methods.

The influence of the number of fusion iterations on predictive performance is illustrated in Fig. 8 for the three test cells A05, B03, and C21. Across all cells and fusion configurations, a consistent pattern is observed. An initial adjustment phase occurs within the first few fusion iterations, followed by either a shallow performance plateau or a gradual degradation in accuracy as the number of iterations increases.

For cell A05, the single-expert model at zero fusion iterations already provides strong baseline performance (RMSE $\approx 5.7 \times 10^{-2}$, $R^2 \approx 0.92$). All fusion configurations follow a monotonic degradation trend as the number of fusion iterations increases, with RMSE gradually rising toward 6.1×10^{-2} and R^2 decreasing toward 0.91. Differences between pairwise and three-expert fusion remain marginal across all fusion iterations.

For cell B03, the single-expert baseline achieves RMSE of approximately 1.86×10^{-2} with R^2 of about 0.82. As fusion progresses, all multi-expert configurations exhibit a moderate increase in RMSE and a corresponding decrease in R^2. Among the tested strategies, pairwise fusion of experts B and C yields the most favourable trade-off, with RMSE reaching around 2.0×10^{-2} and R^2 remaining close to 0.78. In contrast, three-expert fusion consistently results in higher RMSE values and lower R^2 scores across fusion iterations compared with pairwise fusion.

A similar behaviour is observed for cell C21. The initial fusion state (RMSE $\approx 3.4 \times 10^{-2}$, $R^2 \approx 0.80$) is followed by a gradual increase in error with additional fusion iterations. Pairwise fusion with expert A remains closest to the baseline, while three-expert fusion exhibits the largest degradation in both RMSE and R^2.

Table 6 reports the corresponding SOH prediction performance for the individual expert models, the uniform ensemble, and the CIGAR fusion strategies. Based on the convergence behaviour observed in Fig. 8, the number of fusion iterations was fixed to 30 for all fusion configurations.

The single-expert models provide competitive baseline performance on their respective cells. In contrast, the uniform ensemble exhibits substantial degradation across all test cells, characterised by increased RMSE values and negative R^2 scores. The applied CIGAR-based fusion consistently outperforms the uniform ensemble across all cells, while remaining below the corresponding single-expert baselines. Among the tested configurations, pairwise fusion yields the most favourable performance, whereas the inclusion of a third expert does not result in systematic improvement and is often associated with increased error.

Representative SOH profiles obtained from the single-expert models and the CIGAR-based fusion are shown in Fig. 9, together with the corresponding ground-truth SOH trajectories for reference.

Across all cells, the fused predictions preserve the dominant SOH degradation trends in the ground truth, including both gradual capacity fade and sharper transition regions. Importantly, while fusion does not uniformly

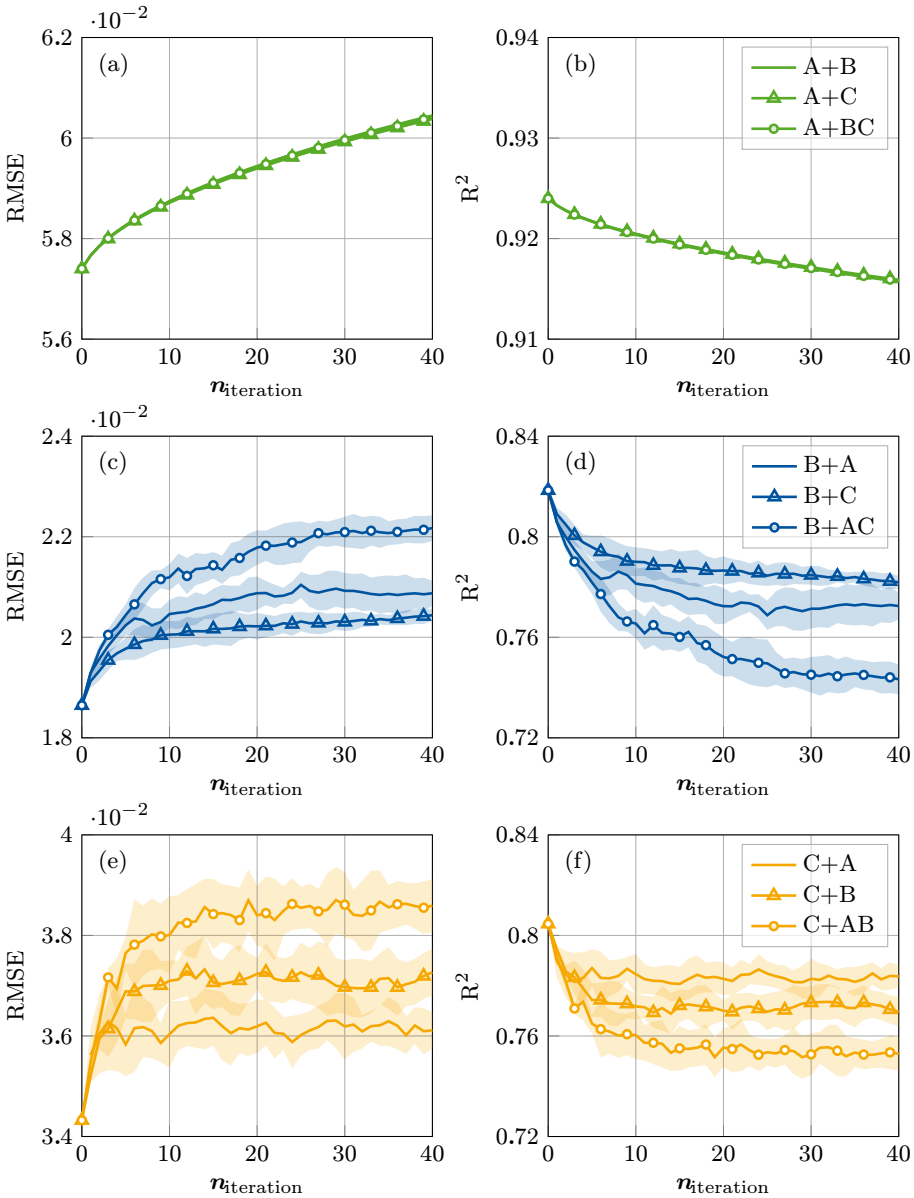

Fig. 8 Influence of the number of fusion iterations on predictive performance using the CIGAR fusion. Results are shown for three test cells: A05 (top row), B03 (middle row), and C21 (bottom row). Columns report (**a**, **c**, **e**) RMSE, and (**b**, **d**, **f**) R^2 as functions of the number of fusion iterations. Solid lines denote the mean performance across runs (μ), while shaded regions indicate one standard deviation ($\pm 1\sigma$). Different curves correspond to distinct fusion strategies, as indicated in the legends

reduce pointwise prediction error relative to the strongest single expert, the resulting SOH trajectories remain comparable.

5 Discussion

The results demonstrate that uniform ensemble averaging across heterogeneous experts leads to severe performance degradation, confirming that simple averaging is not an appropriate strategy for knowledge exchange in this setting. Across all test cases, despite a slight degradation relative to the best single expert, CIGAR consistently outperforms the uniform ensemble. This behaviour indicates that CIGAR is able to combine predictions from heterogeneous experts.

The favourable performance observed for pairwise fusion between cells B03 and C21 can be attributed to their similar cell chemistries and pouch-cell configurations. This observation highlights an important limitation

Table 6 SOH prediction performance on the test cells, reported as mean ± standard deviation over 10 independent runs, with the number of fusion iterations fixed to 30

Method	Test cell	RMSE↓	R^2 ↑
Initial A	A05	0.0574	0.9240
Initial B	B03	0.0186	0.8186
Initial C	C21	0.0343	0.8047
Uniform ensemble	A05	0.2104	−0.0212
Uniform ensemble	B03	0.0478	−0.1946
Uniform ensemble	C21	0.0965	−0.5444
CIGAR (A+B)	A05	0.0600 ± 0.00002	0.9170 ± 0.0001
CIGAR (A+C)	A05	0.0599 ± 0.00002	0.9172 ± 0.0001
CIGAR (A+BC)	A05	0.0600 ± 0.00001	0.9171 ± 0.00002
CIGAR (B+A)	B03	0.0209 ± 0.0004	0.7713 ± 0.0078
CIGAR (B+C)	B03	0.0203 ± 0.0002	0.7847 ± 0.0038
CIGAR (B+AC)	B03	0.0221 ± 0.0003	0.7451 ± 0.0071
CIGAR (C+A)	C21	0.0362 ± 0.0007	0.7827 ± 0.0080
CIGAR (C+B)	C21	0.0370 ± 0.0005	0.7733 ± 0.0066
CIGAR (C+AB)	C21	0.0386 ± 0.0007	0.7527 ± 0.0084

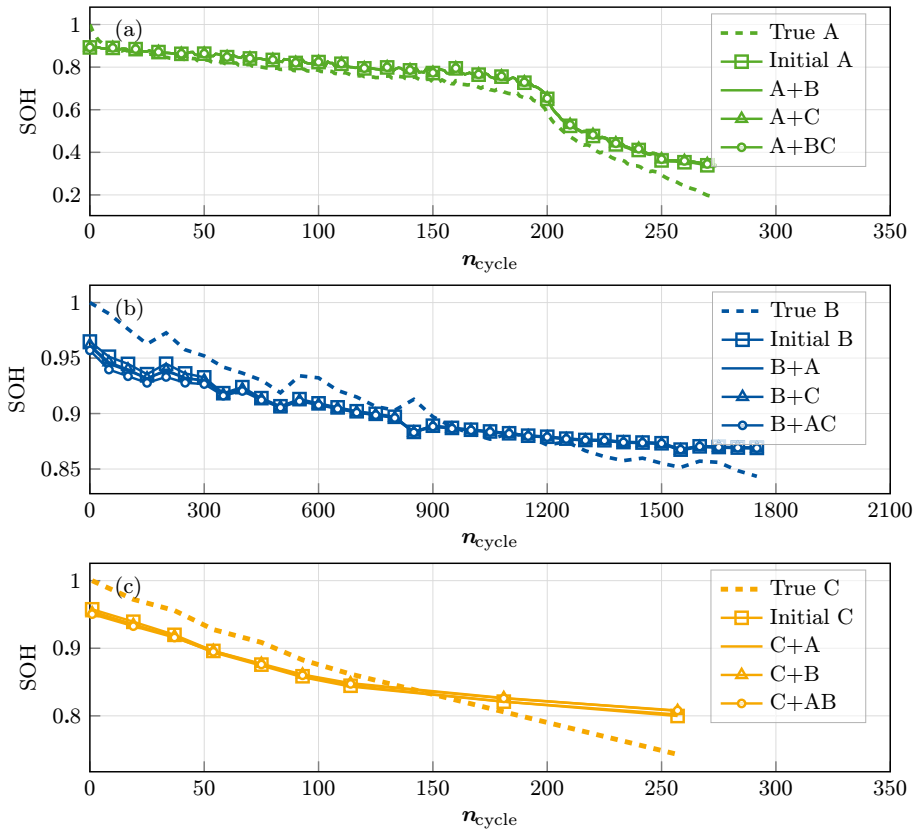

Fig. 9 SOH degradation profiles as a function of cycle number for cells (**a**) A05, (**b**) B03, and (**c**) C21, illustrating the ground truth (dashed), the initial single-expert prediction, and various fusion strategies combining experts A, B, and C

of CIGAR-based fusion, mainly that its effectiveness decreases as heterogeneity between participating experts increases. The reduced performance observed when incorporating a third, less compatible expert provides further evidence that excessive heterogeneity can outweigh the potential benefits of collective inference.

Analysis of the fusion process across iterations shows that the largest adaptation in predictive performance occurs during the initial fusion iterations. This behaviour is inherent to the CIGAR framework, which operates by iteratively merging model outputs rather than retraining internal representations. After the initial adjustment

phase, further fusion iterations no longer improve performance, indicating that the fusion process should be stopped early.

Despite the limited numerical improvements over the best-performing single expert, the fused predictions generated by CIGAR preserve SOH degradation trends. In particular, the resulting SOH trajectories remain consistent with expected ageing behaviour, supporting the applicability of the proposed approach for battery health monitoring tasks.

Overall, the results indicate that CIGAR-based collective inference is most appropriate when applied to distributed models trained on EIS data from cells with compatible chemistries and operating conditions. Under these conditions, CIGAR enables controlled knowledge exchange without direct data sharing, while avoiding the performance degradation associated with uniform ensemble averaging.

6 Conclusion

This work investigates gradient-based collective inference for collaborative SOH prediction using CNN–BiLSTM models trained on three distinct EIS battery datasets. The CIGAR algorithm is employed to enable collaboration between independently trained models without exchanging raw data or model parameters.

The results show that CIGAR can support knowledge exchange between experts when applied under appropriate conditions. In particular, effective fusion was observed when combining models trained on datasets with compatible cell chemistries and cell types, and when limiting the number of fusion iterations. In contrast, uniform ensemble averaging leads to degraded performance in the presence of dataset heterogeneity.

Overall, the findings clarify the conditions under which gradient-based collective inference enables knowledge exchange for SOH monitoring. While the proposed approach does not outperform strong single-expert models, it preserves SOH degradation profiles and avoids the severe performance losses associated with the uniform ensemble method.

Future work should focus on automated expert selection, adaptive stopping criteria, and extending the framework to larger and more diverse battery populations to improve robustness and scalability.

Acknowledgements This work was supported by the project “OptiPro” (grant number: 03XP0364B) as part of the competence cluster “InZePro” funded by the Federal Ministry of Education and Research in Germany (BMBF). This work was also supported by the KIOptiPack project (www.ki-hub-kunststoffverpackungen.de) funded by the German Federal Ministry of Research, Technology and Space (BMFTR) under grant number 033KI115.

Author contributions S. Olbrych: Conceptualisation, Software, Investigation, Project administration, Writing – Original Draft Preparation, Writing – Review & Editing. Z. X. Tung: Software, Investigation, Writing – Original Draft Preparation. S. Celik: Investigation (EIS measurements), Expert consultation, Writing – Review & Editing. Remaining authors: Review, Supervision, Expert Consultation.

Funding Open Access funding enabled and organized by Projekt DEAL.

Data availability The data that support the findings of this study are available from the corresponding authors upon reasonable request.

Declarations

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

References

1. Akbar K, Zou Y, Awais Q et al (2022) A machine learning-based robust state of health (SOH) prediction model for electric vehicle batteries. Electronics 11(8):1216. https://doi.org/10.3390/electronics11081216
2. Blömeke A, Kappelhoff O, Wasylowski D, Ringbeck F, Sauer DU (2024) Open source online electrochemical impedance spectroscopy data analytics tool. J Power Sources 615. https://doi.org/10.1016/j.jpowsour.2024.235049
3. Daniel N, Stoyanov S, Bailey C, et al (2021) Review of Fusion Prognostics for Lithium-Ion Batteries - Current State and Future Challenges. In: 2021 44th International Spring Seminar on Electronics Technology (ISSE). IEEE, p 1–8. https://doi.org/10.1109/isse51996.2021.9467644
4. Faraji-Niri M, Rashid M, Sansom J et al (2023) Accelerated state of health estimation of second life lithium-ion batteries via electrochemical impedance spectroscopy tests and machine learning techniques. J Energy Storage 58:106295. https://doi.org/10.1016/j.est.2022.106295
5. Fong C, Tyler M (2021) Machine learning predictions as regression covariates. Polit Anal 29(4):467–484. https://doi.org/10.1017/pan.2020.38
6. Géron A (2022) Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow, 3rd edn. O'Reilly Media, Sebastopol, CA
7. Gifford CM (2009) Collective Machine Learning: Team Learning and Classification in Multi-Agent Systems. Phd thesis, University of Kansas
8. Guo F, Huang G, Zhang W et al (2023) State of health estimation method for lithium batteries based on electrochemical impedance spectroscopy and pseudo-image feature extraction. Measurement 220:113412. https://doi.org/10.1016/j.measurement.2023.113412
9. Hoang M, Hoang N, Low BKH, et al (2019a) Collective Model Fusion for Multiple Black-Box Experts. In: Chaudhuri K, Salakhutdinov R (eds) Proceedings of the 36th International Conference on Machine Learning, pp 2742–2750, https://proceedings.mlr.press/v97/hoang19a.html
10. Hoang TN, Hoang QM, Low KH et al (2019) Collective Online Learning of Gaussian Processes in Massive Multi-Agent Systems. Proc AAAI Conf Artif Intell 33(01):7850–7857. https://doi.org/10.1609/aaai.v33i01.33017850
11. Jiang B, Zhu J, Wang X et al (2022) A comparative study of different features extracted from electrochemical impedance spectroscopy in state of health estimation for lithium-ion batteries. Appl Energy 322:119502. https://doi.org/10.1016/j.apenergy.2022.119502
12. Kim S, Choi YY, Choi JI (2022) Impedance-based capacity estimation for lithium-ion batteries using generative adversarial network. Appl Energy 308:118317. https://doi.org/10.1016/j.apenergy.2021.118317
13. Li C, Yang L, Li Q et al (2024) SOH estimation method for lithium-ion batteries based on an improved equivalent circuit model via electrochemical impedance spectroscopy. J Energy Storage 86:111167. https://doi.org/10.1016/j.est.2024.111167
14. Li D, Yang D, Li L et al (2022) Electrochemical impedance spectroscopy based on the state of health estimation for lithium-ion batteries. Energies 15(18):6665. https://doi.org/10.3390/en15186665
15. Li W, Peng Y, Zhang M, et al (2023) Deep Model Fusion: A Survey. https://doi.org/10.48550/ARXIV.2309.15698
16. Lin D, Liu Y, Cui Y (2017) Reviving the lithium metal anode for high-energy batteries. Nat Nanotechnol 12(3):194–206. https://doi.org/10.1038/nnano.2017.16
17. Lipu MH, Hannan MA, Hussain A et al (2018) A review of state of health and remaining useful life estimation methods for lithium-ion battery in electric vehicles: challenges and recommendations. J Clean Prod 205:115–133. https://doi.org/10.1016/j.jclepro.2018.09.065
18. Liu D, Luo Y, Liu J et al (2014) Lithium-ion battery remaining useful life estimation based on fusion nonlinear degradation AR model and RPF algorithm. Neural Comput Appl 25(3–4):557–572. https://doi.org/10.1007/s00521-013-1520-x
19. Liu Z, Sun Y, Li Y, et al (2024) Lithium-ion battery health prognosis via electrochemical impedance spectroscopy using CNN-BiLSTM model. Journal of Materials Informatics 4(2). https://doi.org/10.20517/jmi.2024.09
20. Mohtat, Peyman, Siegel, Jason B., Stefanopoulou, Anna G. (2021) Uofm pouch cell voltage and expansion cyclic aging dataset [data set]. https://doi.org/10.7302/7TW1-KC35
21. Momma T, Matsunaga M, Mukoyama D et al (2012) Ac impedance analysis of lithium ion battery under temperature control. J Power Sources 216:304–307. https://doi.org/10.1016/j.jpowsour.2012.05.095
22. Obregon J, Han YR, Ho CW et al (2023) Convolutional autoencoder-based SOH estimation of lithium-ion batteries using electrochemical impedance spectroscopy. J Energy Storage 60:106680. https://doi.org/10.1016/j.est.2023.106680

23. Raschka S (2020) Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning. https://doi.org/10.48550/ARXIV.1811.12808
24. Rastegarpanah A, Hathaway J, Ahmeid M et al (2020) A rapid neural network–based state of health estimation scheme for screening of end of life electric vehicle batteries. Proc Inst Mech Eng I J Syst Control Eng 235(3):330–346. https://doi.org/10.1177/0959651820953254
25. Rastegarpanah A, Hathaway J, Stolkin R (2021) Rapid model-free state of health estimation for end-of-first-life electric vehicle batteries using impedance spectroscopy. Energies 14(9):2597. https://doi.org/10.3390/en14092597
26. Reiter C, Wassiliadis N, Wildfeuer L, et al (2018) Range Extension of Electric Vehicles through Improved Battery Capacity Utilization: Potentials, Risks and Strategies. In: 2018 21st International Conference on Intelligent Transportation Systems (ITSC), p 321–326, https://doi.org/10.1109/itsc.2018.8569455
27. Schindler S, Danzer MA (2017) Influence of cell design on impedance characteristics of cylindrical lithium-ion cells: a model-based assessment from electrode to cell level. J Energy Storage 12:157–166. https://doi.org/10.1016/j.est.2017.05.002
28. Sharma S, Panwar AK, Tripathi M (2020) Storage technologies for electric vehicles. Journal of Traffic and Transportation Engineering (English Edition) 7(3):340–361. https://doi.org/10.1016/j.jtte.2020.04.004
29. Tang Q, Zhang Y, Zhang Y et al (2020) Identifying degradation patterns of lithium ion batteries from impedance spectroscopy using machine learning. Nat Commun 11(1):1706. https://doi.org/10.1038/s41467-020-15235-7
30. Yahya A, Li A, Kalakrishnan M, et al (2017) Collective robot reinforcement learning with distributed asynchronous guided policy search. In: 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp 79–86, https://doi.org/10.1109/IROS.2017.8202141
31. Zhang Q, Huang CG, Li H et al (2022) Electrochemical impedance spectroscopy based state-of-health estimation for lithium-ion battery considering temperature and state-of-charge effect. IEEE Trans Transp Electrif 8(4):4633–4645. https://doi.org/10.1109/TTE.2022.3160021
32. Zhang Y, Tang Q, Zhang Y, et al (2020b) Identifying degradation patterns of lithium ion batteries from impedance spectroscopy using machine learning. Zenodo [data set], https://doi.org/10.5281/ZENODO.3633835

Authors and Affiliations

Sylwia Olbrych[1] · Zi Xuan Tung[2] · Sehriban Celik[3] · Hans Aoyang Zhou[1] · Anas Abdelrazeq[1] · Dirk Uwe Sauer[3] · Robert H. Schmitt[1,4]

✉ Sylwia Olbrych
sylwia.olbrych@wzl-iqs.rwth-aachen.de

✉ Sehriban Celik
sehriban.celik@isea.rwth-aachen.de

[1] Chair of Intelligence in Quality Sensing (IQS), WZL - RWTH Aachen University, Aachen 52074, Germany

[2] RWTH Aachen University, Aachen 52062, Germany

[3] Institute for Power Electronics and Electrical Drives (ISEA), Chair for Electrochemical Energy Conversion and Storage Systems (ESS), RWTH Aachen University, Aachen 52074, Germany

[4] Fraunhofer Institute for Production Technology IPT, Aachen 52074, Germany