

# Transfer learning from synthetic data for open-circuit voltage curve reconstruction and state of health estimation of lithium-ion batteries from partial charging segments

Tobias Hofmann<sup>a,b,\*</sup>, Jacob Hamar<sup>b</sup>, Bastian Mager<sup>b</sup>, Simon Erhard<sup>b</sup>, Jan Philipp Schmidt<sup>a</sup>

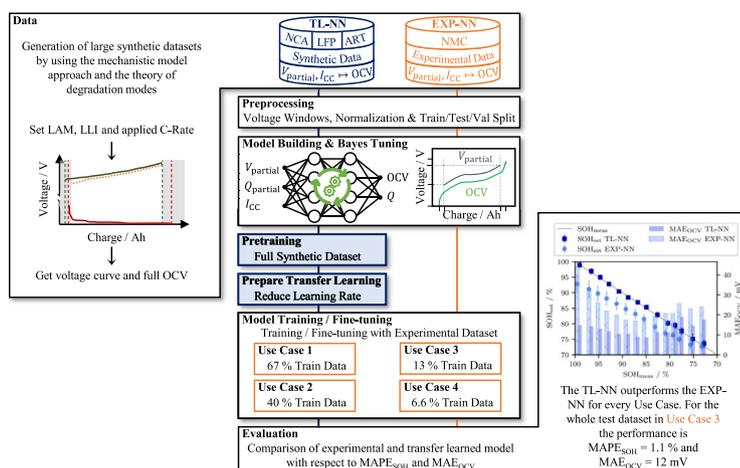
<sup>a</sup> Chair of Systems Engineering for Electrical Energy Storage, University of Bayreuth, Weierstraße 26, Bayreuth, 95448, Bavaria, Germany

<sup>b</sup> BMW Group, Petuelring 130, Munich, 80809, Bavaria, Germany

## HIGHLIGHTS

- A TCN-LSTM model for full OCV reconstruction is introduced.
- Transfer learning from synthetic data is explored for four use cases.
- Transfer learning outperforms conventional deep learning.
- Transfer learning is applied from two other battery datasets.
- Benefits and limitations of transfer learning are explored.

## GRAPHICAL ABSTRACT



## ARTICLE INFO

### Keywords:

Lithium-ion battery  
State of health estimation  
Transfer learning  
OCV curve  
Partial charging  
Synthetic data

## ABSTRACT

Data-driven models for battery state estimation require extensive experimental training data, which may not be available or suitable for specific tasks like open-circuit voltage (OCV) reconstruction and subsequent state of health (SOH) estimation. This study addresses this issue by developing a transfer-learning-based OCV reconstruction model using a temporal convolutional long short-term memory (TCN-LSTM) network trained on synthetic data from an automotive nickel cobalt aluminium oxide (NCA) cell generated through a mechanistic model approach. The data consists of voltage curves at constant temperature, C-rates between C/30 to 1C, and a SOH-range from 70% to 100%. The model is refined via Bayesian optimization and then applied to four use cases with reduced experimental nickel manganese cobalt oxide (NMC) cell training data for higher use cases. The TL models' performances are compared with models trained solely on experimental data, focusing on different C-rates and voltage windows. The results demonstrate that the OCV reconstruction mean absolute error (MAE) within the average battery electric vehicle (BEV) home charging window (30% to 85% state of charge (SOC)) is less than 22 mV for the first three use cases across all C-rates. The SOH estimated from the reconstructed OCV exhibits a mean absolute percentage error (MAPE) below 2.2% for these cases. The study

\* Corresponding author at: Chair of Systems Engineering for Electrical Energy Storage, University of Bayreuth, Weierstraße 26, Bayreuth, 95448, Bavaria, Germany.

E-mail addresses: [tobias1.hofmann@uni-bayreuth.de](mailto:tobias1.hofmann@uni-bayreuth.de), [tobias.th.hofmann@bmw.de](mailto:tobias.th.hofmann@bmw.de) (T. Hofmann).

<https://doi.org/10.1016/j.egyai.2024.100382>

Received 8 April 2024; Received in revised form 14 May 2024; Accepted 2 June 2024

Available online 6 June 2024

2666-5468/© 2024 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

further investigates the impact of the source domain on TL by incorporating two additional synthetic datasets, a lithium iron phosphate (LFP) cell and an entirely artificial, non-existing, cell, showing that solely the shifting and scaling of gradient changes in the charging curve suffice to transfer knowledge, even between different cell chemistries. A key limitation with respect to extrapolation capability is identified and evidenced in our fourth use case, where the absence of such comprehensive data hindered the TL process.

## 1. Introduction

With newer generations of battery electric vehicles (BEVs), more computing resources – either directly in the car or accessible via the backend – accelerate the development of state estimation algorithms. The state of health (SOH) plays an important role because it determines the available energy and power over lifetime. It is mostly defined as the available capacity in relation to its capacity at begin of life (BOL) [1]. Additionally the open-circuit voltage (OCV) is an important characteristic of the lithium-ion battery which is crucial for accurate state of charge (SOC) estimation, algorithms concerning safety, healthy fast-charging and a prolonged lifetime [2,3]. Available methods [4,5] use the pristine open-circuit potential (OCP) curves to reconstruct the aged OCV curve. These reconstruction methods, however, simplify the degradation of the OCV based on the mechanistic model approach [6] and require constant current measurements with very low C-rates [4] or relaxed voltage points [5].

Data-driven methods can bridge this gap by relying on empirical datasets and thus are promising to eliminate the need for specific input measurements. They are capable of discovering data patterns leading to qualitative features for SOH estimation and OCV reconstruction. Especially the combination of deep learning (DL) and explicit input measurements was successful for accurate SOH estimation: Zhang et al. [7] compared their combined incremental capacity analysis (ICA) and broad learning system algorithm with other available data-driven methods. They proved that highly correlating health features from the IC-curve allowed their method to outperform other algorithms and reached SOH mean absolute percentage errors (MAPEs) below 0.42% for three different batteries. In a subsequent study [8] they identified the difficulties incorporated by varying charging manners and working conditions for real-world BEVs. Hence, they introduced a method which solely relies on the raw measurement data from the constant voltage phase, i.e., the gradually decreasing current. Via their proposed double correlation analysis several health features are extracted from that raw sensor data to finally estimate the SOH with a gated recurrent unit. Again, they reached excellent results with SOH MAPEs below 0.48%. Recently, Zhang et al. [9] focused on the available input data from partial charging segments during the constant current phase. With health indicators derived from their newly defined incremental energy per SOC curve, they have trained a bidirectional long short-term memory neural network (LSTM) to estimate the SOH. With random partial charging events at different C-rates (0.5C: 11% to 36%, 0.3C: 23% to 75%, 0.2C: 35% to 82%, 0.1C: 15% to 53%) they have reached SOH mean absolute errors (MAEs) below 0.43%. Dubarry et al. [10] explored the usage of synthetic data for model training and subsequent degradation mode (DM) estimation within photovoltaic systems. They used a mechanistic model to generate voltage curves with varying aging paths and trained several data-driven models on these data. They have validated the method with synthetic data from over 10000 different degradation paths and reached DM root mean squared errors (RMSEs) of 2.75%.

In recent years, partial charging data has also emerged as a suitable input for OCV reconstruction via DL. Tian et al. [11] introduced a convolutional neural network (CNN) that processes the charge amount from parts of the daily charging voltage curves at 1 C to estimate the stoichiometries and hence the OCV curve of aged lithium-ion batteries. With usage of these raw segments they reached an OCV reconstruction RMSE below 15 mV and an SOH estimation error below 1.0%. In the

work by Ruan et al. [12] they have used a CNN as a quick analysis tool to estimate DMs from OCV curves. They developed the method fully with synthetic data but have shown the real-world applicability by using partial OCV segments from an experimental battery. With a voltage segment from 3.4 V to 4.18 V they have estimated the DMs with a RMSE of 1.75%. These DMs were further used to reconstruct the full OCV with a RMSE of 21.05 mV. Guo et al. [13] used a similar approach where they required partial charging segments at 1C to estimate partial OCV curves with a LSTM. The partial OCV was fed into an optimizer which reconstructed the full OCV with a MAE below 20 mV and estimated the SOH with a MAPE below 1.3%. In contrast to the CNN developed by Tian et al. [11], the proposed LSTM by Guo et al. [13] struggled to directly interpret raw data and, hence, first calculated health-indicators from the charging curves.

Both, CNNs and LSTMs, have specific benefits for battery state estimation from time-series data: CNNs, which were initially proposed for computer vision [14], do not require additional feature engineering and can process raw data segments. LSTMs, however, are beneficial for time-series data processing due to their ability to handle sequential data including internal feedback loops. Specifically in the last years, temporal convolutional neural networks (TCNs) [15], which combine the benefits from CNNs and LSTMs, have gained attraction in the field of battery state estimation [16–20]. These new types of neural networks (NNs) outperform former architectures, eliminate the need for cumbersome preprocessing and are optimized for time-series data [16]. Without preprocessing, Bockrath et al. [16] have fed raw sensor data from partial discharging segments into a TCN network and reached an overall SOH RMSE for unseen test data of 1.0%. Their model outperformed several other NN architectures. In their study the middle and upper SOC-range was more suitable for SOH estimation. Li et al. [17] compared the performance of different NN architectures for SOH estimation using the same dataset as Bockrath et al. [16]. They proved that TCN-LSTM networks increase the accuracy by more than 16% compared to TCN, LSTM and CNN-LSTM networks, i.e., the MAE decreased from 3.11% (TCN), 3.38% (LSTM), 2.78% (CNN-LSTM) to 2.5% (TCN-LSTM).

DL, however, is accompanied by challenges such as data dependence and limited training data availability [21]. DL models typically require vast amounts of data to effectively capture complex data patterns. The training data must encompass a significant portion of the data space to enable the model to interpolate between different locations. To address this issue, transfer learning (TL) has been introduced, eliminating the requirement for training and test data to be independent and identically distributed [21]. In the developed TL model by Sahoo et al. [22], they mapped features from the voltage curve to the SOH which lead to SOH estimation errors below 2% for unseen data. A two-layered feed-forward neural network (FNN) was pretrained with an experimental dataset. Finally, TL was applied by freezing the first layer and fine-tuning the second layer with other public datasets. Zou et al. [23] followed the same model-based TL approach with a different dataset and architecture: They used experimental battery aging data from the public NASA degradation dataset. Source and target domain, however, were selected from this dataset which was divided into groups based on varying usage conditions. Their work, hence, evaluated TL within the same cell chemistry to new degradation paths. Their NN processed features from the constant current charging voltage curve to the CNN input layer. A LSTM layer interpreted the output of the CNN layer and passed its output through a FNN layer which produced the SOH value. The freezing mechanism was limited to the LSTM layer and the

amount of training data from the target domain was varied during fine-tuning. They showed that with limited training data from the target domain, TL models outperform stand-alone NNs. In contrast, Shen et al. [24] took a different approach by omitting the freezing mechanism and instead focused on retuning multiple models to construct an ensemble model. Their base model consisted of five CNN and three FNN layers, which were utilized to process the raw voltage, current, and charge time-series data from partial constant current charging curves for estimating the SOH. Rather than fine-tuning a single model, the authors replicated the architecture to create several new models, each of which was subsequently fine-tuned using different subsets of the training data domain. Tian et al. [25] finally explored the transferable knowledge between synthetic and experimental data by creating a CNN for SOH estimation. They have pretrained the model with synthetic data from a voltage capacity model, simulating the experimental cell. Fine-tuning was carried out with 300 mV partial charging segments. They repeated the experiment for three public available datasets and reached SOH RMSEs of 1.00%, 3.13% and 0.52% for the Oxford, CALCE and Tongji dataset [25], respectively. Zhou et al. [26] were the first to apply TL for full OCV reconstruction from daily partial charging data. With a generative DL model, they successfully applied TL from different cell chemistries with SOH RMSEs of 0.47% and 2.73% for nickel cobalt aluminium oxide (NCA) and lithium iron phosphate (LFP) cells, respectively.

A critical limitation identified in these models is their reliance on extensive experimental training datasets. While a multitude of public datasets are available, they may not always align with specific research objectives or provide the ideal conditions for model training. Zhou et al. [26] highlighted the potential for TL in OCV reconstruction, a topic that has received limited attention in the existing literature.

The present study introduces a novel approach for OCV reconstruction and SOH estimation utilizing a TL framework with a TCN-LSTM architecture. This approach is distinctive in its exclusive use of synthetic data from a different cell derived from partial voltage charging curves to train the model. Synthetic data for an automotive NCA cell are generated using a mechanistic model approach, with the dataset partitioned according to different voltage windows. The use of synthetic data allows fast data generation and detailed analytics. A base model is trained on this dataset and optimized using Bayesian optimization techniques. Subsequently, TL is applied to four distinct use cases (UCs), each characterized by varying quantities of experimental, measured nickel manganese cobalt oxide (NMC) cell data, to evaluate the efficacy and constraints of TL. The performance of TL models is benchmarked against models trained solely on experimental data for each UC, thereby assessing the influence of TL on model outcomes. The evaluation focuses on the effects of varying C-rates and voltage windows. Notably, we determine from a field data analysis that the average BEV home charging pattern is compatible with the proposed algorithm. Additionally, the study explores the influence of the source domain on TL by incorporating two alternative synthetic datasets for base model training, one based on a LFP cell and the other entirely artificial. This allows to interpret the hidden features in voltage trajectories necessary to reconstruct the OCV. We give evidence that the shift of plateaus in the voltage curve is the main information captured by DL and especially TL models.

To our knowledge, this is the first study to employ synthetic data as the source domain for TL between varying cells in the context of OCV reconstruction, as well as the first to investigate TL from wholly artificial cells. This approach provides novel insights into the learning dynamics of DL models.

## 2. Theory

The proposed NN estimates the full OCV curve from constant current partial charging segments. The estimated SOH is finally calculated from that OCV curve. The NN, hence, only processes the raw voltage

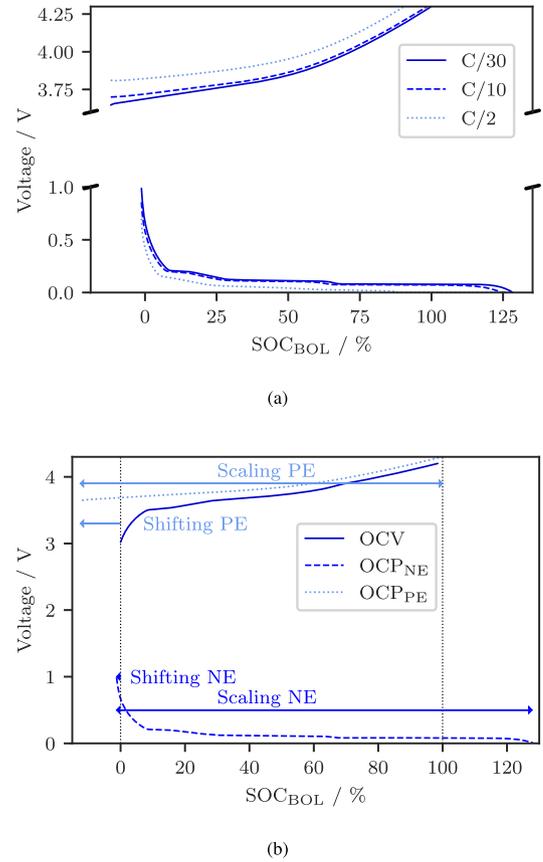


Fig. 1. Fundamentals of the mechanistic model approach from Dubarry et al. [6]. The C-rate refers to the full-cell capacity measurable during a constant current charge. (a) To generate constant current charging curves at different C-rates the model relies on half-cell potential measurements at different C-rates. (b) The user has to define the degradation parameters, which results in a scaling and shifting of the half-cell potential curves and finally output the voltage curve as the difference between the positive and the negative half-cell potential curve.

curve, the capacity throughput during this partial charging event and the applied C-rate. The utilization of machine learning (ML) techniques for battery modeling and state estimation, however, is currently constrained by the substantial data prerequisites. The application of TL on synthetic data presents an alternative approach to circumvent the necessity for expensive aging experiments. This section, hence, is mandatory to follow the subsequent method section and introduces the synthetic data generation model and elucidates the fundamentals of TL. In the remainder of this paper, we refer to the following definition for the SOH: The SOH is defined as the measured charge capacity at a low C-rate, e.g., C/25, relative to its nominal capacity at BOL. The amount of charge is measured between lower and upper voltage limits.

$$\text{SOH} = \frac{C_{\text{meas}, C/25}}{C_N} \quad (1)$$

We further simplify the OCV definition to include pseudo-OCV measurements at C-rates as low as C/25.

### 2.1. Mechanistic model for battery data generation

The mechanistic model approach, originally proposed by Dubarry et al. [6], posits that the OCV of a lithium-ion battery throughout its lifespan can be determined by the disparity between the half-cell OCP curves of the positive and negative electrode [27,28]. The shape of the OCV is primarily influenced by a relative shift and scaling of the OCPs. If the degradation-induced impedance rise is disregarded,

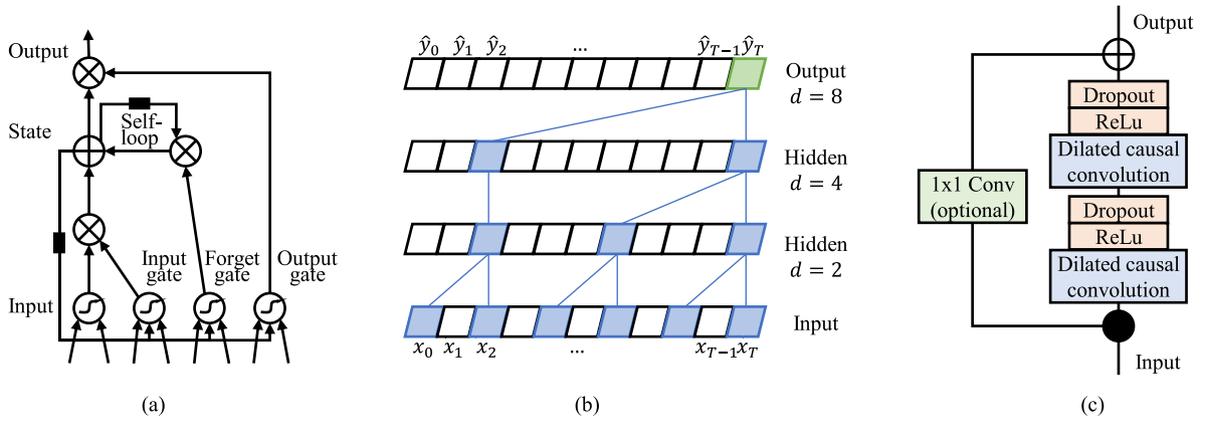


Fig. 2. The main building blocks of LSTM ((a) adopted from Goodfellow et al. [29]) and TCN ((b), (c) adopted from Bai et al. [34] and Bockrath et al. [16]) networks. (a) Memory cell of a LSTM network. (b) Example for one dilated causal convolution within a residual block for a kernel size of 2 and dilation factors  $d = 8, 4, 2$ . (c) TCN residual block with the optional  $1 \times 1$  convolutional in case of different dimensions between input and output. ReLU is the most common activation function which only passes positive input, i.e.,  $f_{\text{ReLU}}(x) = \max(0, x)$ .

this concept can be extended to voltage responses to applied current, allowing for the reconstruction of voltage curves at higher C-rates using the measured half-cell potential curves at corresponding C-rates [6]. To generate data, the mechanistic model employs pristine half-cell potential curves at various C-rates, as illustrated in Fig. 1(a). In the mechanistic model, the degradation rate is determined solely by the user-defined shifting and scaling parameters, as illustrated in Fig. 1(b). This fundamental theory allows the simulation and generation of mass synthetic data with predefined DMs and SOH based on real measurements. For a more comprehensive understanding of these alignment parameters, readers are encouraged to refer to the works of Dubarry et al. [6] or Hofmann et al. [5]. In our example (Fig. 1), the model generates the OCV curve from the OCP measurements, along with the SOH. This implies that the process of constructing the OCV curve from Fig. 1(b) can also be applied to C-rates of C/10 and C/2. Consequently, the model can be utilized to produce constant current voltage curves at any given C-rate, provided that pristine half-cell potential measurements are available and overpotentials are linear.

## 2.2. Neural networks for battery state of health estimation

With the advances in ML, more network architectures arise to estimate the SOH from battery cycling data. The simplest architecture in DL is the FNN, which consists of multiple neurons and direct connections between layers. It can be described as a function approximator  $y = f(x; \theta)$ , where  $y$  represents the approximated value (e.g., SOH),  $x$  represents arbitrary features extracted from battery aging data (e.g., voltage level or C-rate), and  $\theta$  represents the hyperparameters [29,30]. The FNN does not have any feedback loops and information flows only in a forward direction.

In contrast, if the network includes feedback loops between neurons or layers, it is known as a recurrent neural network (RNN). RNN are particularly suitable for handling sequential data, such as time-series data [29].

LSTMs, as in Fig. 2(a), form a specific type of RNN. First introduced by Hochreiter and Schmidhuber [31], they have gained immense popularity in the field of battery diagnosis [32,33]. These memory cells (Fig. 2(a)) are compact, self-regulating networks that possess input, output, and forget gates. By means of an internal feedback loop, the LSTM mechanism adjusts the weights based on the temporal history. Consequently, the network determines both the type and quantity of memory to be stored or discarded [29].

Similar to RNNs, CNNs have recently gained increased attention in the field due to their ability to process grid-like data, i.e., multiple time-series data streams, without preprocessing [29,35,36]. Filters convolve over the  $n$ -dimensional input and pass the results to the next layer.

These filters allow the inclusion of the present time-step, as well as historical data [35,36]. Some adaptations [17,23,37] leverage the feature extraction power of CNNs sequential to the feature processing ability of LSTMs to build accurate SOH estimation models.

TCNs were proposed to couple the benefits from CNNs and LSTMs: Lea et al. [15] introduced the novel TCN initially for action segmentation and detection. This special type of CNN uses causal convolution to produce output with the same length as the input, similar to RNNs [16]. TCNs outperform RNNs in terms of training efficiency, model size, model accuracy and do not require data preprocessing [16]. Bai et al. [34] evaluate the architecture for a variety of use cases against common RNN architectures and propose TCNs as a natural starting point for sequence modeling, i.e.,  $\hat{y}_0, \dots, \hat{y}_T = f(x_0, \dots, x_T)$ . Fig. 2(b) and (c) illustrate the main building blocks of a TCN. Causal convolution leads to an output with same length as the input and avoids leakage from future information  $x_{N+1}$  to the past  $\hat{y}_N$  [16]. The output at time  $\hat{y}_T$  is only convolved with input data at time  $T$  and earlier  $x_0, \dots, x_T$ . In contrast to CNNs, which use centered kernels, TCNs employ kernels where the rightmost value of the kernel is the considered time [16,34]. When it comes to DL or large receptive fields, causal convolution leads to many additional weights and stabilization problems. Dilated convolution (Fig. 2(b)) solves this problem by dilating the kernel of size  $k$  by the dilation factor  $d$ . The dilated convolution operation  $F()$  on the element  $s$  is

$$F(s) = (x *_d f)(s) = \sum_{i=0}^{k-1} f(i) \cdot x_{s-d \cdot i} \quad (2)$$

for a 1D-sequence input  $x \in \mathbb{R}^n$  and the filter  $f : \{0, \dots, k-1\} \rightarrow \mathbb{R}$ . The final receptive field per layer is then calculated by

$$R = 2^l(k-1) \quad (3)$$

where  $l$  is the number of layers. Hence, this allows the TCN to process exponentially large receptive field which is especially helpful for efficient processing of long time-series data [16,34]. Finally, a residual block (Fig. 2(c)) allows to learn modifications to the identity mapping instead of the entire transformation [16]. One residual block consists of stacked dilated causal convolutions, ReLU and dropout blocks. An optional  $1 \times 1$  convolution processes the input in case of differing dimensions between input and output. The output  $o$  of a residual block is the applied transformation  $\mathcal{F}$  added to the input  $x$  of the block [34]:

$$o = \text{Activation}(x + \mathcal{F}(x)) \quad (4)$$

### 2.3. Fundamentals of transfer learning

In the domain of TL it is essential to establish a mathematical understanding of its principles. While DL models encounter challenges related to data dependence and insufficient training data [21], TL aims for a solution. This approach eliminates the necessity for training and test data to adhere to the assumption of being independent and identically distributed.

To understand the mathematical definition of TL we require some basic annotations and refer to the publication by Tan et al. [21] and Weber et al. [38]: A domain  $\mathcal{D} = \{\chi, P(X)\}$  is defined by the feature space  $\chi$  and the edge probability distribution  $P(X)$ , where  $X = \{x_1, \dots, x_n\} \in \chi$ . A task  $\mathcal{T} = \{y, f(x)\}$  consists of the label space  $y$  and the target prediction function  $f(x)$ , which can be also understood as the conditional probability function  $P(y | x)$ .

**Definition 2.1 (Transfer Learning).** TL aims to enhance the performance of the predictive function  $f_{\mathcal{T}}()$  for the main learning task  $\mathcal{T}_t$  by leveraging latent knowledge discovered from the supplementary learning task  $\mathcal{T}_s$  and its associated dataset  $D_s$ , where  $D_s \neq D_t$  and/or  $\mathcal{T}_s \neq \mathcal{T}_t$ . Typically, the size of  $D_s$  is significantly larger than  $D_t$ , i.e.,  $N_s \gg N_t$ .

In most cases – and also in this paper – TL refers to *domain adaption* (DA), i.e.,  $D_s \neq D_t$  [38]. In the literature [21,38], TL for time-series data consists of three categories:

**Instance-based** TL assumes that instances from the source and target domain are similar. Hence, selected or reweighted samples from the source domain are included in the TL process.

**Feature-based** TL transfers both domains into the same feature representation, i.e., uses the same features for both datasets.

**Model-based** TL is the most common form of TL and reuses a pretrained model in the source domain for target model building. In this context, several possibilities exist: Parameter-based TL refers to only reusing the learned parameters. Ensemble TL describes the stacking of a pretrained model with a blank model for the target domain. Model controlled TL changes the objective function for retraining. Finally, adversarial-based TL utilize generative adversarial networks [21,38].

In the context of battery state estimation and OCV reconstruction mostly model-based TL exists [22–24,26]. Hence, we elaborate model-based TL for SOH estimation: A straightforward way to deploy TL for SOH estimation is partial freezing and fine-tuning, a strategy used within parameter-based TL, as in the work by Sahoo et al. [22]. Hence, an initial model is trained on the source domain, while just a subset of the pretrained model is retrained on the target domain. The remaining frozen parts of the retrained model do not change their parameters, i.e., the weights of the neurons, during retraining. The retraining process, also known as fine-tuning, typically involves reducing the learning rate and/or limiting the number of epochs to allow only slight modifications of the model parameters. Both mechanism, freezing and fine-tuning, hinder the TL model from catastrophic forgetting [38].

In summary, most researchers choose model-based TL approaches with partial freezing and/or fine-tuning to deploy their TL battery SOH estimation models.

### 2.4. Performance metrics

In our work we always refer to the MAPE for SOH estimation and to the MAE for OCV reconstruction accuracy. The MAPE and MAE are defined as follows:

$$\text{MAPE} = \frac{1}{N} \sum_{i=1}^N \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (5)$$

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (6)$$

In Eqs. (5) and (6),  $y_i$  represents the ground truth, i.e., the measured OCV or SOH, whereas  $\hat{y}_i$  corresponds to the estimation;  $N$  is the total number of samples. When considering the OCV reconstruction error  $\text{MAE}_{\text{OCV}}$ , it is crucial that both  $y_i$  and  $\hat{y}_i$  must be vectors of the same length. Consequently, if the model outputs an OCV that is longer or shorter than the ground truth, the error is only calculated at the overlapping section. Eqs. (5) and (6) can be modified for our study as follows:

$$\text{MAPE}_{\text{SOH}} = \frac{1}{N} \sum_{i=1}^N \left| \frac{\text{SOH}_{\text{meas},i} - \text{SOH}_{\text{est},i}}{\text{SOH}_{\text{meas},i}} \right| \quad (7)$$

$$\text{MAE}_{\text{OCV}} = \frac{1}{N} \sum_{i=1}^N |\text{OCV}_{\text{meas},i} - \text{OCV}_{\text{est},i}| \quad (8)$$

## 3. Materials and method

The efficacy of ML techniques is heavily reliant on the underlying data and its distribution. A significant limitation of data-driven approaches for SOH estimation is the requirement for numerous aging experiments across multiple aging paths. Performing such experiments is both expensive and time-consuming. TL from synthetic data presents a promising avenue for developing ML methods with limited data. Given the crucial role of data sources, this paper presents them first, followed by a detailed exposition of the investigated TL approach.

### 3.1. Data

This work utilizes synthetic data created by a mechanistic model tool. The simulation toolbox relies solely on the measured pristine half-cell potential curves at various C-rates. Instead of modeling the physico-chemical degradation mechanisms in numerical equations it uses the simplified theory of DMs [6]. In this work, we consider loss of active material at the anode ( $\text{LAM}_{\text{NE}}$ ), loss of active material at the cathode ( $\text{LAM}_{\text{PE}}$ ) and loss of lithium inventory (LLI). Similar to the work of Dubarry et al. [6] the used approach shifts and scales the pristine half-cell potential curves to set specific DMs. One simulation always includes the charging profile for the set C-rate and also the OCV, because the half-cell potential curves are captured for various C-rates. For a detailed description of DMs and their utilization for voltage curve generation we refer to a previous publication [5].

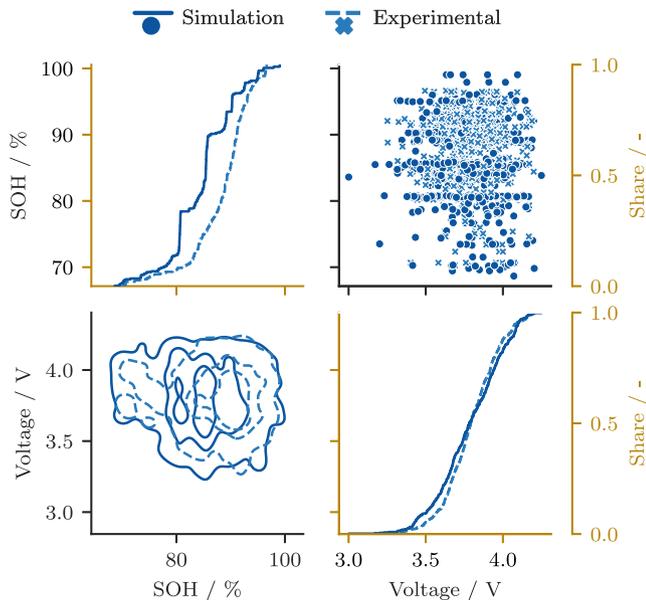
The simulation data is based on pristine half-cell potential measurements of a prismatic automotive NCA-graphite cell with 70.2 Ah at an ambient temperature of 25 °C. The layer thicknesses are  $t_{\text{NE}} = 107 \mu\text{m}$  and  $t_{\text{PE}} = 62.5 \mu\text{m}$ , respectively. The initial loading ratio (LR) and SOC<sub>PE</sub> offset (OFS) are 0.95 and 10.6%. For the concept of LR and OFS we refer to the work of Dubarry et al. [6]. The simulation is carried out by varying all three DMs:  $\text{LAM}_{\text{NE}}$  and  $\text{LAM}_{\text{PE}}$  are varied between 0% to 15% in steps of 3.75%. LLI is varied in steps of 5% between 0% to 20%. By setting this generic degradation parameters, in total  $5 \cdot 5 \cdot 5 = 125$  voltage curves exist per C-rate. Per voltage curve, the corresponding OCV curve is available, too. All voltage and OCV curves are cut to the range from 3 V to 4.25 V.

The experimental dataset, comprising 10 commercial INR18650-MJ1 NMC-silicone/graphite cells (LR: 0.90, OFS: 11%, silicon mass ratio in anode between 1wt.% to 5wt.%), is taken from Schmitt et al. [4] and preprocessed to only include the constant current sections. The investigated cells were partitioned into five different aging studies and hence different aging paths, as listed in Table 1. Between degradation phases, all cells underwent an OCV measurement and a charging rate test. The OCV measurement refers to a constant current charging measurement with C/30 and the charging rate test data, used within this work, refers to the constant current measurements at 0.26C  $\approx$  C/3.85, C/2 and 1C. The C/3.85 charge is equivalent to a constant-power 11 kW charge for a 42 kWh storage with a 96s36p configuration, to evaluate the applicability of the proposed algorithm by Schmitt et al. [4] for real use cases. Again, the SOH refers to the OCV measurement. Due to the

**Table 1**

Aging conditions of the experimental dataset, taken from Schmitt et al. [4]. All degradation phases ran at an ambient temperature of 25 °C.

Cell number	Aging conditions
A1, A2	2.5 V–4.2 V cycling
B1, B2	2.5 V–4.0 V cycling
C1, C2	3.6 V–4.2 V cycling
D1, D2	WLTP cycling
E1, E2	Calendar aging at 3.7 V



**Fig. 3.** Visualization of the synthetic NCA and the experimental NMC dataset. The upper right figure shows the scatter plot and the lower left figure visualizes the kernel densities. Both refer to the first y-axis on the left. The second y-axis on the right side refers to the cumulative distribution plots in the diagonal, in golden color. All voltage curves at any given C-rate are illustrated. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

increasing minimal voltage at begin of charge, all samples are cut to the voltage-range between 3.24 V to 4.2 V.

The synthetic dataset is visualized together with the experimental dataset in Fig. 3. For both datasets, the SOH for a given voltage curve refers to the accumulated charge between the lower and upper voltage limit during an OCV measurement. In both datasets, all samples with a SOH below 70% are excluded due to our definition of the end of life (EOL) at 70% SOH. While the SOH range seems evenly distributed for the experimental dataset, the predefined degradation states are visible in the synthetic SOH trajectory. Due to the similar cathode material of both datasets, the voltage distribution looks similar. This is, as well, visible in the kernel densities and the scatter plot.

Fig. 4 provides a more comprehensive analysis of the data. As shown in Fig. 4(a), the synthetic dataset encompasses charging trajectories ranging from C/15, C/10, C/6, C/3, C/2 and 1C, while the OCV is established through a constant current-charging process with C/25. Conversely, the experimental dataset in Fig. 4(b) includes constant current-charging events at C/3.85, C/2, and 1C. In contrast to the synthetic dataset, the OCV measurement is obtained from a C/30 charge. The voltage curve exhibits an increase for higher currents primarily due to the ohmic overpotentials, which display a direct correlation with the applied current. Consequently, the upper voltage limits are reached earlier, resulting in a shorter constant current-phase. In this study, we exclude the CV-phase from all calculations and base our SOH and SOC calculations solely on the constant current-phase. The SOH always refers to the full charge amount between the voltage limits in the OCV measurement. Fig. 4(c) and (d) illustrate pristine

and aged OCV curves and the correlation between SOH and the OCV trajectory. The observations can be categorized into two fundamental aspects of battery aging: capacity fade and impedance rise. Capacity fade is directly evident in Fig. 4(c) and (d) by the earlier reached cutoff voltage. In the synthetic dataset, the capacity is modeled by the DMs. Impedance rise is only evident in the experimental datasets, visualized in Fig. 4(d). Impedance rise also contributes to reaching the cutoff voltage earlier. Impedance rise over lifetime is not included in the synthetic dataset because it aggravates the clear interpretation of plateau shifting and scaled OCPs. Hence, in this paper we focus on the effect of capacity drop and TL, while the influence of impedance rise will be discussed in future work.

Both datasets were captured at 25 °C to focus on the clear effect of transferred knowledge from different cell chemistries, C-rates and voltage windows. The real-world applicability of the method, however, is still given due to negligible temperature dependencies of the voltage and OCV curve between temperatures from 20 °C to 40 °C [39].

To further evaluate the data requirements for model fine-tuning and especially the urge to cover most of the dataset, several UCs are defined. The described UCs divide the experimental dataset for fine-tuning into differing data segments, which vary in their difficulty based on content and amount of training data. This separation aims to explore the boundaries of NNs and especially TL for battery state estimation. Fig. 5 visualizes the training, validation and test dataset per UC. The UCs for TL can be differentiated based on two factors: Firstly, they vary in terms of the included charging curves at different C-rates. Secondly, they differ in terms of the included degradation path. The data split comprises four distinct UCs, namely UC1 (Fig. 5(a)), UC2 (Fig. 5(b)), UC3 (Fig. 5(c)), and UC4 (Fig. 5(d)). These UCs cover different boundaries of the training dataset. UC1 to UC3 encompass the entire SOH range, including the lowest and highest C-rate. In contrast, UC4 only includes the entire SOH range for one specific C-rate. Approximately 67% of the entire experimental dataset is covered by UC1, while UC2 covers 40%, UC3 covers 13%, and UC4 covers only 6.6%. In UC1, all cells and charging curves are used for training, except for the curves at C/2. UC2 further complicates the TL process by excluding aging path B, as well as one cell each from aging path C and D from the training dataset. Similarly, UC3 reduces the training dataset even further by excluding the cells from aging path D and E completely. The most challenging UC4 solely includes the charging curves at C/3.85 from cell A2 and C2 for training.

### 3.2. Default charging window

To evaluate the appropriate charging window for analysis, we collect customer field data comprising 1.9 million BEV AC-charging events conducted at residential locations. The dataset includes the corresponding SOC windows of those charging events. Based on the findings depicted in Fig. 6, the most frequently utilized charging window is identified, which serves as the focus of our study. The distribution of start-SOC values is remarkably uniform, with the mean occurring at approximately 49%.

To establish a default start-SOC, we opt for the minimum start-SOC that encompasses at least 25% of all charging events. Consequently, a start-SOC of 30% is selected. Analysis of the end-SOC values reveals two prominent peaks at 80% and 100% SOC, resulting in a mean value of 85%. Consequently, this value is adopted as the default end-SOC.

In conclusion, we establish a default SOC charging window spanning from 30% to 85%. This window serves as a benchmark for our study and further demonstrates the practicality and applicability of our methodology to real-world charging events at resident locations.

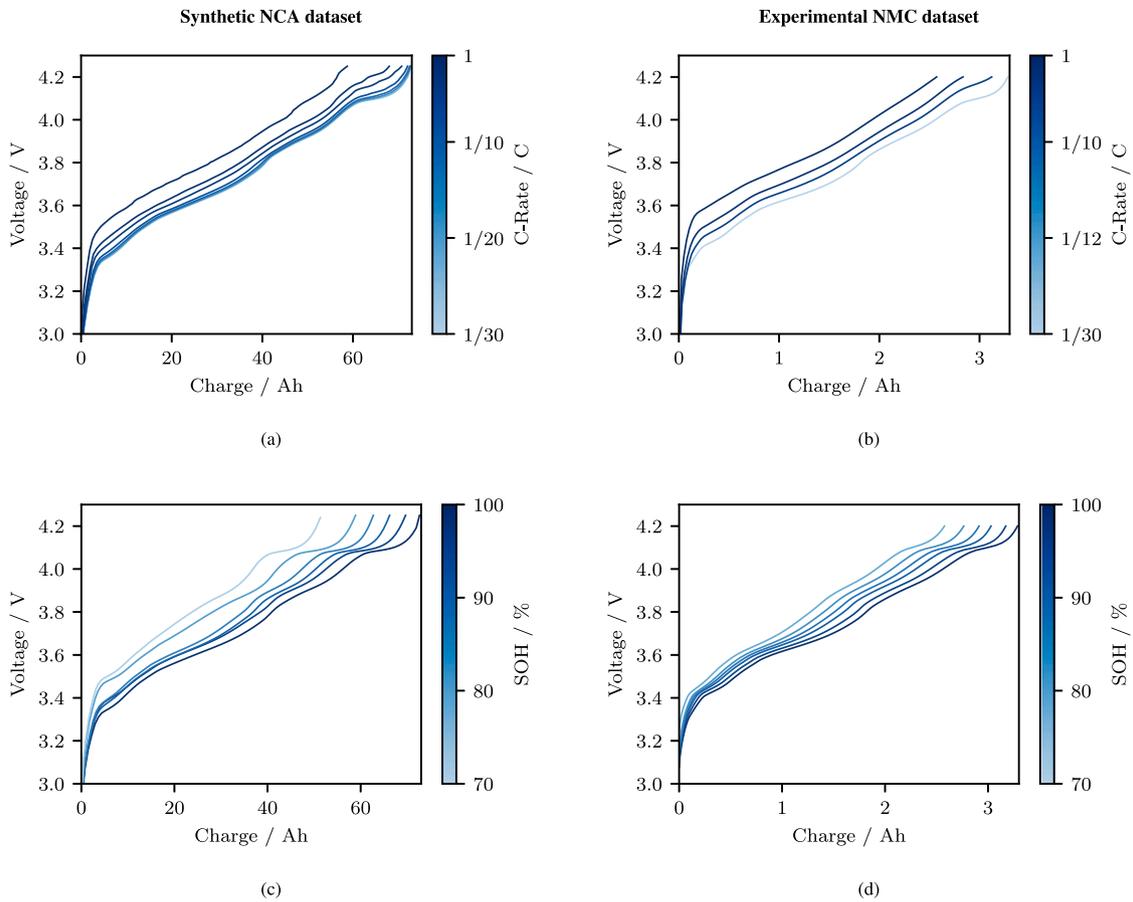


Fig. 4. Influence of C-rate and SOH on the voltage trajectory for the synthetic NCA and the experimental NMC dataset. (a) Influence of C-rate: Synthetic dataset. (b) Influence of C-rate: Experimental dataset. (c) Influence of SOH on OCV: Synthetic dataset. (d) Influence of SOH on OCV: Experimental dataset.

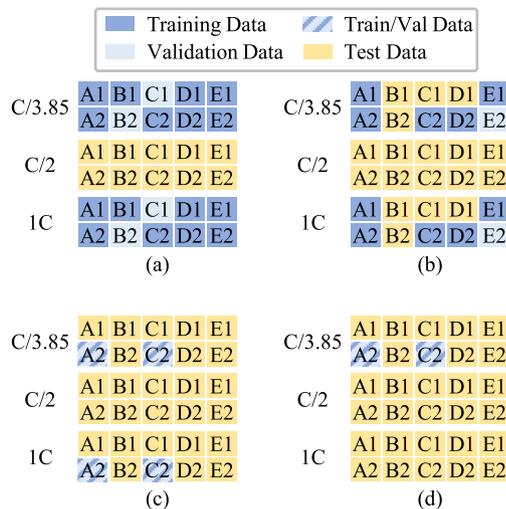


Fig. 5. Definition of different UCs for model comparison: (a) UC1, (b) UC2, (c) UC3 and (d) UC4. The UCs are distinctive in their used curves from different cells and C-rates of the experimental dataset for training data, visualized by different colors. In UC3 and UC4 a random 33% subset of the training data is used for validation. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

### 3.3. Method

The method is summarized in Fig. 7(a), which describes the steps necessary to build the transfer learned neural network (TL-NN) and

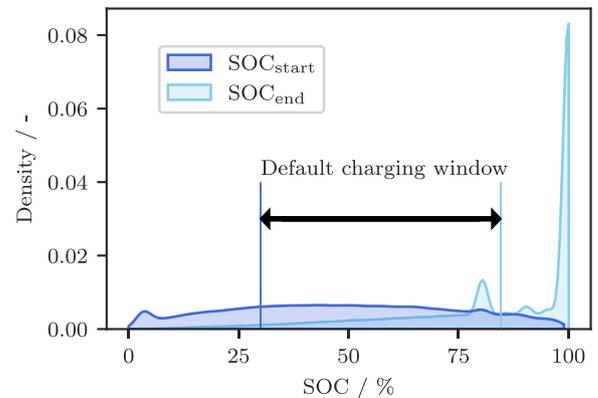


Fig. 6. Start- and end-SOC for 1.9 million BEV AC-charging events. The default partial charging window is selected from the SOC start point with at least 25% share. (In comparison: The average start-SOC is 49%.) The end-SOC is the average for all charging events. The default partial charging window is from 30% to 85%.

the reference EXP-NN. The EXP-NN is only trained and evaluated with the experimental dataset to evaluate the influence of TL. In the main method, the TL-NN is pretrained with the NCA synthetic dataset, hence only referred to as TL-NN. For discussing the limitations and benefits of TL, and the exploration of the underlying transferred knowledge, additional TL-NNs are set up with a LFP and an artificial (ART) cell. These developed models are referred to as TL (LFP) and TL (ART) model, respectively. Both datasets undergo preprocessing, are normalized and split to training, test and validation sub-datasets as indicated

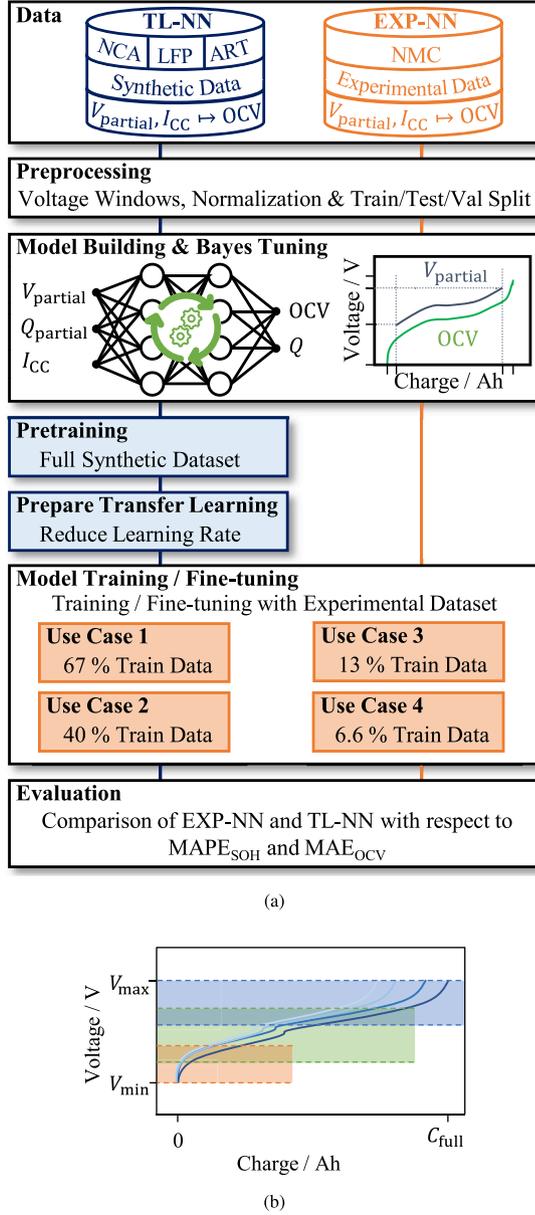


Fig. 7. Description of the TL approach. (a) Overview of the applied method and the differentiation between the TL and the reference EXP-NN path. The main TL path is carried out with the NCA cell. (b) Example for three different partial voltage segments extracted from four different constant current voltage curves.

in Fig. 5. Both models are optimized with Bayes tuning to yield the best architecture. The TL-NN is subsequently pretrained with the full synthetic dataset. The pretrained TL-NN and the not-trained EXP-NN are finally fine-tuned/trained with partial amounts of data from the UCs, defined in Fig. 2. Both models are evaluated with the  $MAE_{\text{SOH}}$  and  $MAE_{\text{OCV}}$ . We choose a similar approach as Bockrath et al. [16] and snipped the full voltage charge curve into multiple partial segments based on voltage boundaries in 200 mV steps, as in Fig. 7(b). In Fig. 7(b)  $C_{\text{full}}$  refers to the charged capacity during a full charge at the given C-rates. The example can be applied to every voltage curve and any C-rate. For the upper voltage segments above 4 V, the step-size is lowered to 100 mV, in order to allow more variation in the final end SOC. Instead of just three partial segments we utilize more steps and extract 19 different voltage windows from one simulated full voltage curve. From the experimental dataset, we extract 15 different voltage windows per full voltage curve. This generates more training data and allows a more

Table 2

Selected voltage windows from the simulation and experimental dataset during preprocessing.

	$V_{\text{min}}$	$V_{\text{max}}$		$V_{\text{min}}$	$V_{\text{max}}$
	3.25 V	3.8 V		3.0 V	3.8 V
	3.4 V	4.0 V		3.2 V	4.0 V
Exp. Data	3.6 V	4.1 V	Sim. Data	3.4 V	4.1 V
	3.8 V	4.2 V		3.6 V	4.25 V
				3.8 V	

detailed analysis of the sensitivity to different voltage segments. Table 2 summarizes the chosen minimum and maximum voltage values  $V_{\text{min}}$ ,  $V_{\text{max}}$  from all selected windows. All possible combinations, e.g.,  $V_{\text{min}} = 3.25$  V and  $V_{\text{max}} = 4.2$  V, are selected. A partial charge window is only extracted if  $V_{\text{min}} < V_{\text{max}}$ . Hence, with this approach, in real-world scenarios arbitrary charging data can be snipped into the respective voltage segments if and only if the start voltage is below 3.6 V and the end voltage is above 3.8 V or the start voltage is below 3.8 V and the end voltage is above 4.0 V.

All compared networks are trained with the same feature set comprising of the raw voltage curve, the partial charge curve (which is defined to start at 0), and the applied C-rate. The dimension of each feature is  $\mathbb{R}^{1 \times 100}$  which requires a downsampling for all data sources. The scalar C-rate is also mapped to a  $\mathbb{R}^{1 \times 100}$  vector to match the dimensions. The whole dataset undergoes normalization to lie in a range between zero and one for further processing. This has been shown to increase NN performance and efficiency [40]. Note that both datasets, the source and target dataset, are normalized individually. During every tuning or training process, the dataset is split into training, validation and test data as shown in Fig. 5. To ensure a representative validation dataset for UC3 and UC4, these datasets are selected as a random 33% subset of the training dataset.

TCNs have been shown to be advantageous for time-series modeling because they work as a type of automated feature extractor from raw data [34]. Many authors [16–20] have, besides that, already proven the ability to employ TCNs for SOH estimation. In our work, we combine TCN layers as a type of feature extractor with subsequent LSTM layers as a type of feature interpreter. Other works [17,23,37] show the mutual support of CNN with LSTM, and even TCN with LSTM. We select fine-tuning as our approach due to the similarity of both datasets. The TCN layer(s) extract similar features for both the source and target domains, while the LSTM layer(s) interpret these features in a similar manner. Hence, only small modifications of the network should be necessary during TL. To accomplish this, we reduce the learning rate and retrain the network for a smaller number of epochs, allowing only minor changes in the network configuration. The proposed method can, of course, be applied to different NN architectures as well. It is important, however, to note that the comparison of these architectures is beyond the scope of this publication.

Before applying TL, the NNs for both the experimental and synthetic datasets are optimized to their respective best performance using Bayesian hyperparameter tuning. Bayesian optimization is a highly effective hyperparameter tuning method as it enables exploration of a large search space with minimal trials [41]. This method employs a surrogate model to minimize the validation loss for different hyperparameter sets. The algorithm prunes unpromising directions and selects the most promising hyperparameter combination for its subsequent trial, thereby requiring only a few trials to achieve optimal performance. Table 3 lists the search space of the hyperparameter tuning. The batch-size is fixed to 128 and the number of epochs is set to a maximum of 500 with early stopping once the validation loss does not decrease for 50 consecutive epochs. Default values are used for all hyperparameters not listed in Table 3. We use the Adam optimizer with the default learning rate of  $1 \times 10^{-3}$ . The last layer is always a LSTM layer with 2 units to guarantee correct output dimensions. To optimize the performance of

**Table 3**

Search space for Bayesian hyperparameter tuning, where  $N$  refers to the layer number. Hence, for every additional TCN or LSTM layer, the respective hyperparameters can be tuned individually.

Hyperparameter	Definition
TCN layers	$\lambda_{\text{TCN}} \in A_{\text{TCN}}$ with $A_{\text{TCN}} := \{\lambda_{\text{TCN}} \in \mathbb{N}; 1 \leq \lambda_{\text{TCN}} \leq 3\}$
Kernel size	$\lambda_{\text{K},N} \in A_{\text{K}}$ with $A_{\text{K}} := \{\lambda_{\text{K}} \in \mathbb{N}; 1 \leq \lambda_{\text{K}} \leq 8\}$
Filter size	$\lambda_{\text{F},N} \in A_{\text{F}}$ with $A_{\text{F}} := \{\lambda_{\text{F}} \in \mathbb{N}; 8 \leq \lambda_{\text{F}} \leq 128\}$
LSTM layers	$\lambda_{\text{LSTM}} \in A_{\text{LSTM}}$ with $A_{\text{LSTM}} := \{\lambda_{\text{LSTM}} \in \mathbb{N}; 0 \leq \lambda_{\text{LSTM}} \leq 2\}$
Units	$\lambda_{\text{U},N} \in A_{\text{U}}$ with $A_{\text{U}} := \{\lambda_{\text{U}} \in \mathbb{N}; 8 \leq \lambda_{\text{U}} \leq 128\}$
Dropout	$\lambda_{\text{D},N} \in A_{\text{D}}$ with $A_{\text{D}} := \{\lambda_{\text{D}} \in \mathbb{D}; 0 \leq \lambda_{\text{D}} \leq 0.5\}$

our model, we explored a large search space. This includes tuning the number of TCN layers  $\lambda_{\text{TCN}}$  (range 1 to 3) with the number of filters  $\lambda_{\text{F},N}$  (range 8 to 128) and kernel size  $\lambda_{\text{K},N}$  (range 1 to 8), as well as the number of LSTM layers  $\lambda_{\text{LSTM}}$  (range 0 to 2) with the number of units  $\lambda_{\text{U},N}$  (range 8 to 128) and dropout  $\lambda_{\text{D},N}$  (range 0 to 0.5) between layers. By considering a wide range of hyperparameters, we aim to find the optimal configuration for our specific task. The optimization runs for 100 trials with the KerasTuner [42] and the goal to minimize the validation loss. To tune the architecture of the reference EXP-NN, we select cells with C-rate values of 1C and C/3.85 for training. Cells B2 and C1 with C-rate C/2 are chosen for validation, while the remaining cells serve as the test dataset to assess the model's generalization ability. This train/test/validation split, optimized for UC1, includes training data from other UCs, enabling the architecture to learn patterns in this data. Fig. 8(a) illustrates the optimized architecture. It consists of an input-TCN layer combined with three LSTM layers and dropout applied between them.

Similar to the EXP-NN, we split the synthetic dataset based on C-rates to create training, testing, and validation datasets.

All curves with C/6 are used for validation, while all curves with C/3 are used for testing, and the remaining data is used for training. The resulting model architecture is depicted in Fig. 8(b). Unlike the EXP-NN, this architecture only includes one TCN layer and no additional LSTM layers besides the output layer. The synthetic dataset is cleaner and does not include any measurement noise, allowing the model to learn data patterns quickly without the need for a complex architecture.

For all four UCs, both the EXP-NN and TL-NNs utilize the same data split, as outlined in Fig. 5: In UC1, cells B2 and C1 from the training dataset are selected as the validation data. In UC2, cell E2 from the training dataset is chosen. For UC3 and UC4, a random subset comprising 33% of the training data is used as the validation dataset to avoid further reduction of the training data space.

The EXP-NN trains in every UC for a maximum of 500 epochs with early stopping once the validation loss does not decrease for 50 consecutive epochs. The training runs for 250 epochs for UC1, 442 epochs for UC2, 500 epochs for UC3 and 471 epochs for UC4.

To achieve optimal learning, the base synthetic model is initially trained with the complete synthetic dataset before TL. A random subset comprising 33% of the data is used for validation during this training phase. The base model successfully achieves  $\text{MAPE}_{\text{SOH}}$  and  $\text{MAE}_{\text{OCV}}$  values below 1% and 10 mV, respectively, for all samples across different C-rates. In contrast to the pretraining phase, the objective of TL is to make marginal modifications to the parameters of the base model. This is accomplished by setting the learning rate to  $1 \times 10^{-6}$  and training for only 200 epochs.

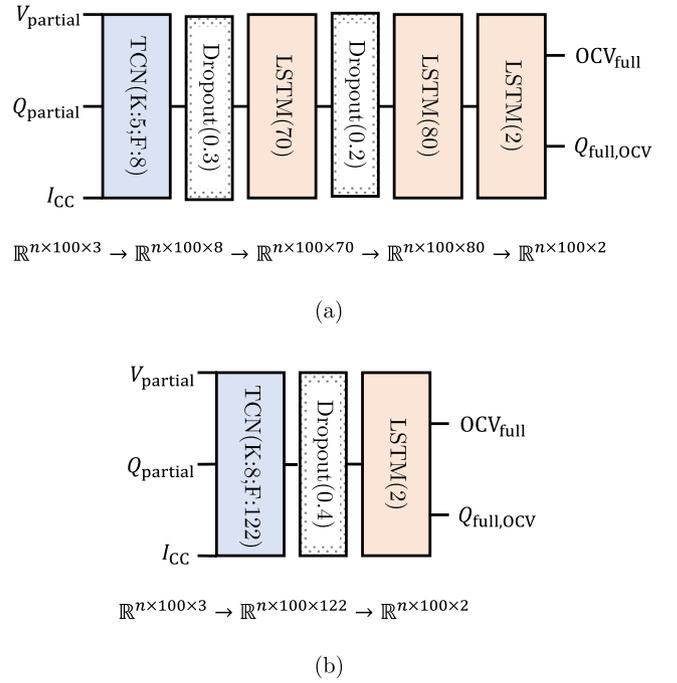


Fig. 8. Final architecture of the tuned (a) EXP-NN and (b) synthetic model / TL-NN and the corresponding dimensions of output data from each layer with  $n$  as the number of samples. In the TCN-layers the first number refers to the kernel size and the second number to the filter size.

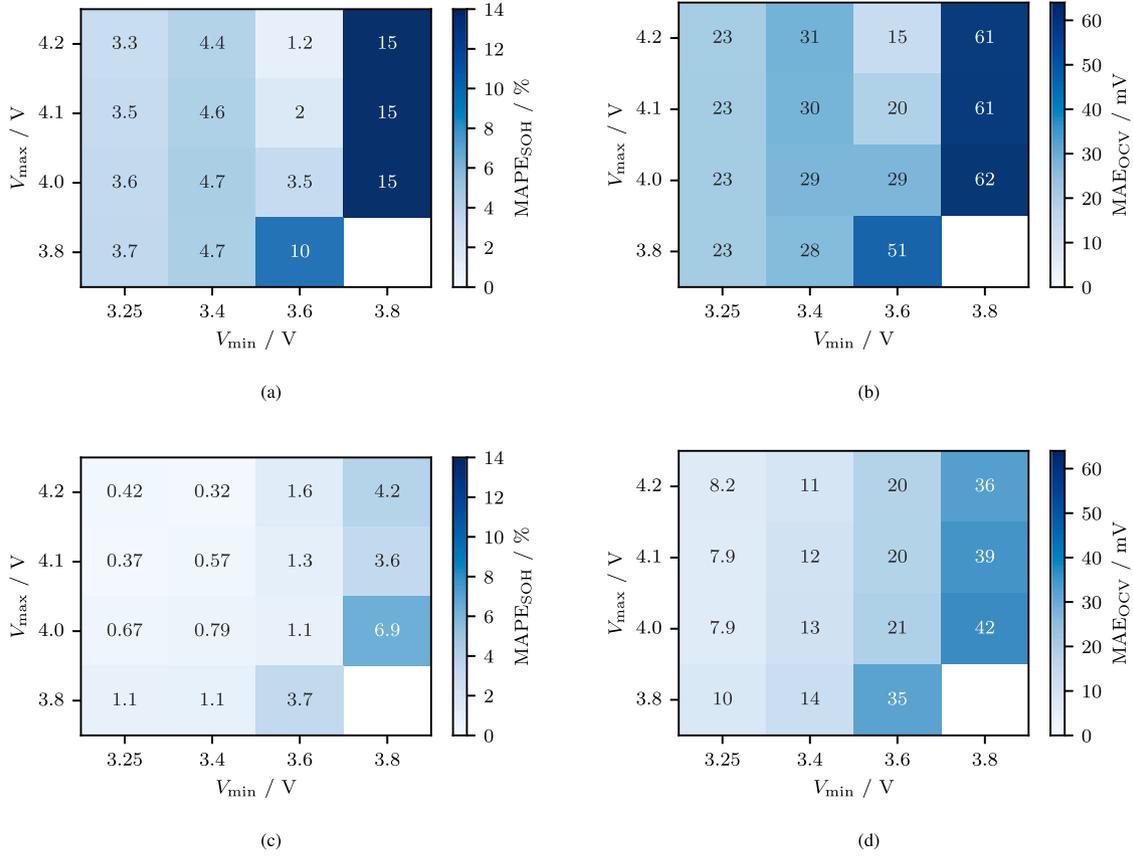
## 4. Results and discussion

The present study aims to evaluate the benefits and limitations of TL, with the UCs defined in Fig. 5 guiding the subsequent discussion. While the EXP-NN is expected to perform comparably to the TL-NN when sufficient training data is available, the TL-NN is expected to outperform the EXP-NN in all other cases. In our approach, the cell chemistries used within the target and source domain are similar. Due to the underlying theory of the mechanistic model approach, the sole information of plateau shifting along degradation should be enough to reconstruct the OCV. Hence, the possibilities of TL from other cell chemistries is discussed in the latter subsection. We even designed a purely artificial synthetic dataset (ART) from self-designed half-cell potential curves to prove this assumption.

### 4.1. Benchmark

The evaluation of the UCs is based on the models derived from the Bayes optimization (Fig. 8). These UCs vary in terms of the amount and content of training data. Table 4 provides a summary of the results, with a focus on the  $\text{MAE}_{\text{OCV}}$  and  $\text{MAE}_{\text{SOH}}$  for reconstruction from partial segments between 3.6 V to 4.1 V. The voltage window is selected because it represents the available charging window from 30% to 85% at BOL, extracted from Fig. 6. Hence, Table 4 gives realistic results for possible applications.

While the more complex EXP-NN still achieves reasonable results for UC1 and UC2, it worsens drastically for the remaining UCs. In a similar manner, the advantage of the simpler TL-NN gets more visible once the training data for fine-tuning gets more and more limited. This, however, also implies the limitations of TL for sparse fine-tuning datasets which do not cover the boundaries of the target domain. For all UCs, the models are evaluated within different partial charging segments.



**Fig. 9.** Results for UC1: (a), (b) EXP-NN, (c), (d) TL-NN. (a) Average SOH error for the test dataset with the reference EXP-NN. (b) Average OCV error for the test dataset with the reference EXP-NN. (c) Average SOH error for the test dataset with the TL-NN. (d) Average OCV error for the test dataset with the TL-NN.

**Table 4**

Performance comparison of the EXP-NN and the TL-NN for different UCs in the voltage window between 3.6 V to 4.1 V. The UCs differ in the amount of training data for the training/fine-tuning process.

		MAPE <sub>SOH</sub>			MAE <sub>OCV</sub>		
		C/3.85	C/2	1C	C/3.85	C/2	1C
UC1	EXP-NN	–	2.0%	–	–	21 mV	–
	TL-NN	–	1.3%	–	–	20 mV	–
UC2	EXP-NN	2.8%	4.9%	2.5%	18 mV	18 mV	15 mV
	TL-NN	0.6%	2.0%	1.1%	8.0 mV	22 mV	10 mV
UC3	EXP-NN	7.2%	2.9%	3.3%	27 mV	22 mV	18 mV
	TL-NN	0.5%	2.2%	0.8%	7.0 mV	20 mV	8.0 mV
UC4	EXP-NN	1.2%	6.7%	20.4%	15 mV	140 mV	280 mV
	TL-NN	0.9%	6.8%	17.0%	7.0 mV	31 mV	25 mV

#### 4.1.1. Use Case 1: Retraining from all cells at two C-rates

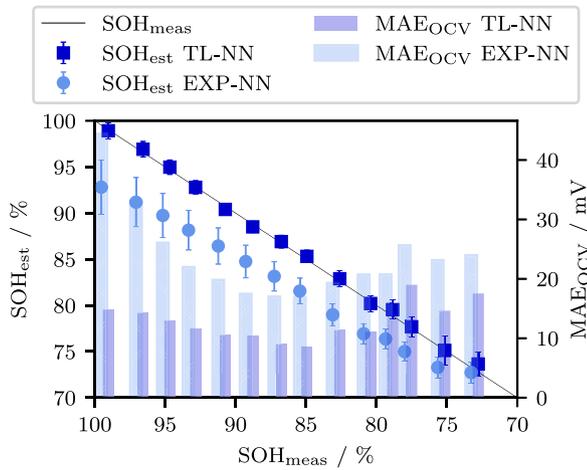
Fig. 9 illustrates the performance of both models across different partial charging segments. It is evident that both models exhibit higher accuracy for wider voltage-ranges. Interestingly, the EXP-NN shows a decrease in performance specifically for the partial charging segments starting at 3.4 V, while the TL-NN does not display this trend. This observation suggests that TL may have been successful in improving the performance of the TL-NN for these segments. The TL-NN consistently outperforms the EXP-NN across all voltage segments, except for the range between 3.6 V to 4.2 V. Notably, it becomes evident that the lower charging segment, starting at 3.25 V or SOC = 0%, is sufficient for reconstructing the OCV with a MAE below 10 mV, even with a lower upper voltage limit. This voltage window, spanning from 3.25 V to 3.8 V, corresponds to a SOC window between 0% and 30% to 50% over the battery's lifetime.

#### 4.1.2. Use Case 2: Retraining from six cells at two C-rates

The results for UC2 are presented in Appendix A.1, Fig. A.14. We train and fine-tune the models using only 40.4% of the experimental dataset, with training data consisting of samples with C-rates of C/3.85 and 1C. The TL-NN outperforms the EXP-NN in terms of SOH and OCV accuracy. Table 4 shows that the TL-NN performs slightly worse by 4 mV for the test dataset at C/2 and in the voltage-range between 3.6 V to 4.1 V. The TL-NN, however, still achieves reasonable results. Notably, the TL-NN's performance for C/2 remains consistent compared to UC1, even with limited training data. On the other hand, the EXP-NN exhibits worsened SOH accuracy with a MAPE of 4.9%. Both models demonstrate a trend of better performance for wider voltage windows. Similar to UC1, we obtain the best results with a low starting voltage of 3.25 V, even with limited voltage windows. However, the accuracy decreases for the upper voltage segment, with an MAE<sub>OCV</sub> above 30 mV. This is mainly due to two main factors: First, the low voltage segments include the steepest gradient at the begin of charge which allows accurate estimation of the anode degradation. Second, the lower voltage segments include more and significant gradient changes, as can be seen in Fig. 4. This effects provide useful information for the developed NNs and hence increase the reconstruction accuracy.

#### 4.1.3. Use Case 3: Retraining from two cells at two C-rates

UC3 further reduces the experimental training data to only 13.2% at two C-rates. Appendix A.2 presents the performance matrix for different voltage windows in Fig. A.15. The TL-NN exhibits nearly the same accuracy as in the previous UCs. Even for the test data at C/2 and between 3.6 V to 4.1 V, as shown in Table 4, the TL-NN performs comparably. The TL-NN outperforms even its own previous realization for UC1, owing to the larger amount of test data from the same C-rate, as revealed in Table 4: The TL-NN performs significantly better



**Fig. 10.** Evaluation of the EXP-NN and the TL-NN for UC3 and the whole test dataset. One data point corresponds to the average estimation within a 2% region and bins them for better comparability. The markers for the SOH estimations indicate the mean value, while the bars correspond to the standard deviation. Similar, the mean OCV reconstruction error is evaluated with the bar plot. To increase comparability between the EXP-NN and TL-NN, the results from the TL-NN are shifted to the left by 0.25%, and the results from the EXP-NN are shifted to the right by 0.25%.

for C/3.85 and 1C. Also the EXP-NN seems to perform better for UC3, but only for lower voltage segments with  $V_{\min} \leq 3.4$  V.

The advantages of TL become even more pronounced when comparing the accuracy along the degradation, as depicted in Fig. 10. The TL-NN shows excellent performance in terms of SOH accuracy with almost no recognizable error and barely any visible standard deviation. For the whole test dataset, i.e., all C-rates and voltage windows, the  $MAPE_{SOH}$  is 1.1% and the  $MAE_{OCV}$  is 12 mV. Also, the OCV reconstruction accuracy stays below 18 mV for every SOH. Below 75% SOH, the standard deviation gets visible with values below 1.6% which is further recognizable by the higher OCV reconstruction error. A correlation between the OCV reconstruction error and the SOH estimation accuracy is thus probable.

Contrary, the EXP-NN underestimates the battery SOH along its lifetime by approximately 5.2% and further inhibits a large standard deviation of up to 2.9%. The  $MAE_{OCV}$  is 24 mV. The EXP-NN performs worse for the upper SOH region with higher SOH standard deviations and a larger OCV reconstruction error. Again, this supports the previously made assumption about an existing correlation between the OCV reconstruction error and the SOH estimation accuracy. In total, the OCV reconstruction accuracy deviates between 17 mV to 47 mV.

Fig. 11 presents the OCV reconstruction from partial voltage segments at C/2 and 1C for different SOHs for the TL-NN at UC3. The figure is divided into three columns with different SOHs and four rows with varying input voltage windows: The low SOC-range ((a)–(c)) spans from 3.25 V to 3.8 V, the mid SOC-range ((d)–(f)) from 3.6 V to 4.1 V, the high SOC-range ((g)–(i)) from 3.8 V to 4.2 V and the full SOC-range ((j)–(l)) from 3.25 V to 4.2 V. Additionally, in every subplot, the  $MAE_{OCV}$  for both C-rates is given in its respective color and position. The TL-NN performs very well for the lower SOC-range for all three degradation states, as can be seen in Fig. 11(a)–(c). The error stays mostly below 12 mV. Solely, the OCV reconstruction with 1C at the lowest SOH (Fig. 11(c)), shows a worse result with a  $MAE_{OCV}$  of 22 mV. With higher SOC-ranges, the accuracy worsens. While the TL-NN reaches acceptable reconstruction results for mid of life (MOL) (Fig. 11(e) and (h)) and EOL (Fig. 11(f) and (i)), the BOL reconstruction (Fig. 11(d) and (g)) from the partial C/2 shows high deviations with a  $MAE_{OCV}$  of 37 mV. A more detailed analysis reveals that the trajectory of the reconstructed OCV shows a good fit but is shifted to the left. On average, the upper voltage segments in Fig. 11(g)–(i) lead to the worst results with OCV

reconstruction errors between 9 mV and 37 mV. Again, for the BOL fit from the partial C/2 voltage curve the trajectory itself looks promising but is shifted to the left. More interesting, the input data from the partial segment at 1C lead to better results than the inputs from the C/2 data. This stands in contrast with recent literature findings about the applicability and limitations of mechanistic OCV reconstruction approaches [4,43–47]. The performance of the TL-NN heavily relies on the training data from the target domain. As long as gradient changes in the charging voltage curve are recognizable, the NN does not seem to be influenced by the C-rate. Similar results were already gathered from evaluations of the base simulation model, where also no C-rate dependency could be found. The full SOC-range (Fig. 11(j)–(l)) leads to similar results as the low SOC-range (Fig. 11(a)–(c)). This is already expected due to the heatmaps investigated from Fig. A.15. All OCVs are reconstructed with sufficient accuracy.

#### 4.1.4. Use Case 4: Retraining from two cells at one C-rate - The limitations of transfer learning

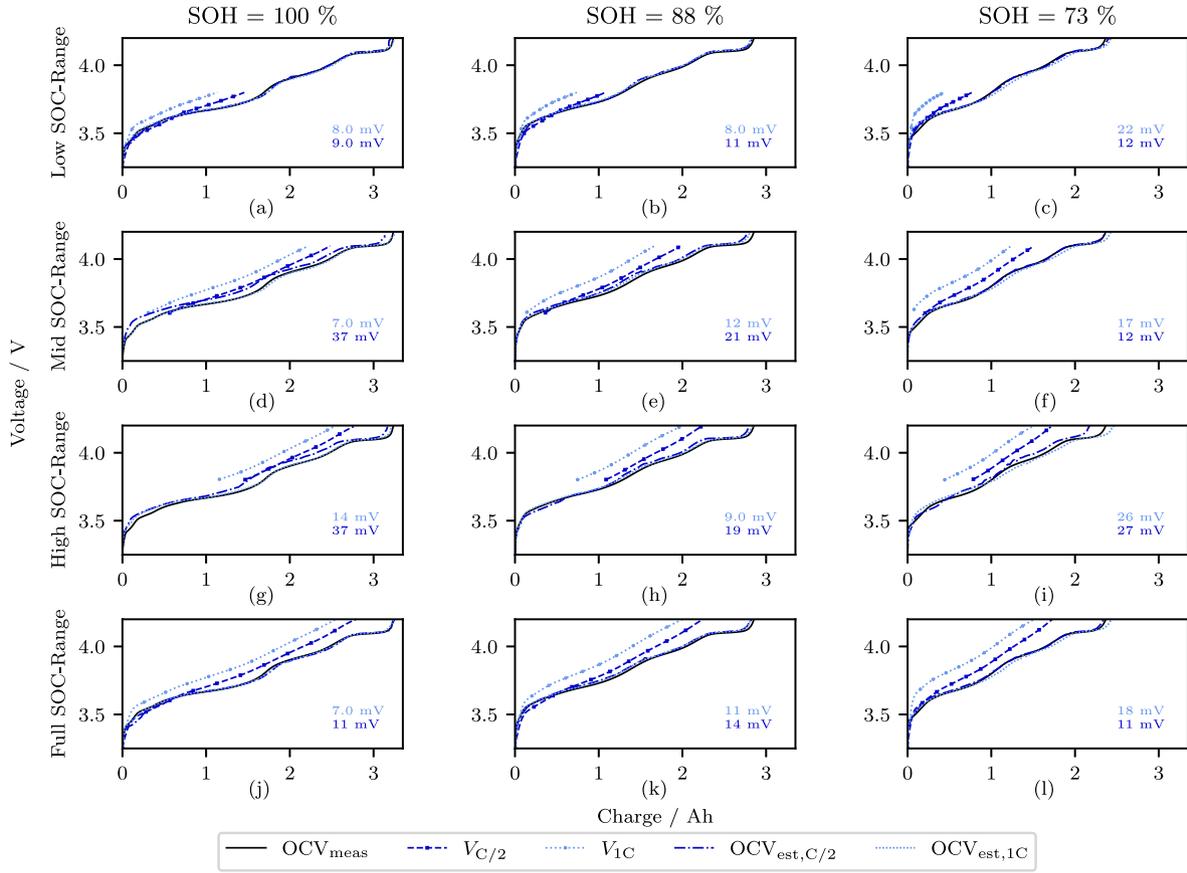
Within UC4 the limitations of TL get visible, as Fig. A.16 in Appendix A.3 shows. As expected, the EXP-NN completely fails to estimate the OCV with a  $MAE_{OCV}$  up to 280 mV for test data at 1C from the voltage window between 3.6 V to 4.1 V due to insufficient training data. In comparison, the TL-NN performs well with an average OCV reconstruction error of 25 mV. Fig. A.16(b) indicates that the estimated OCV curves are shifted to the left or right and hence lead to less accurate results. The OCV reconstruction error can only be evaluated at charge positions where both curves – the measured and reconstructed OCV – are available. This explains the lower  $MAE_{OCV}$ . The previously seen patterns in the accuracy depending on the voltage windows are not visible for the EXP-NN. As can be seen in Fig. A.16 the model achieves the best approximation with errors below 130 mV for the upper voltage segments. Of course, this must be interpreted in the context of the already very high errors and thus should not lead to any conclusions. In contrast, the TL-NN still shows comparable OCV reconstruction dependency as the previous UCs: The OCV is reconstructed with a  $MAE_{OCV}$  below 20 mV for the full and lower voltage segments, while it worsens up to 26 mV for the upper voltage segments.

This indicates that TL is indeed able to transfer knowledge from the source domain to the target domain if, and only if, the fine-tuning process includes the boundaries of the target domain. In our study, these boundaries are samples with C/3.85 and 1C at BOL and EOL. Hence, TL works perfectly for UC3 but suddenly fails for UC4 once the samples with 1C are excluded from the training data, as can be seen in Fig. 5.

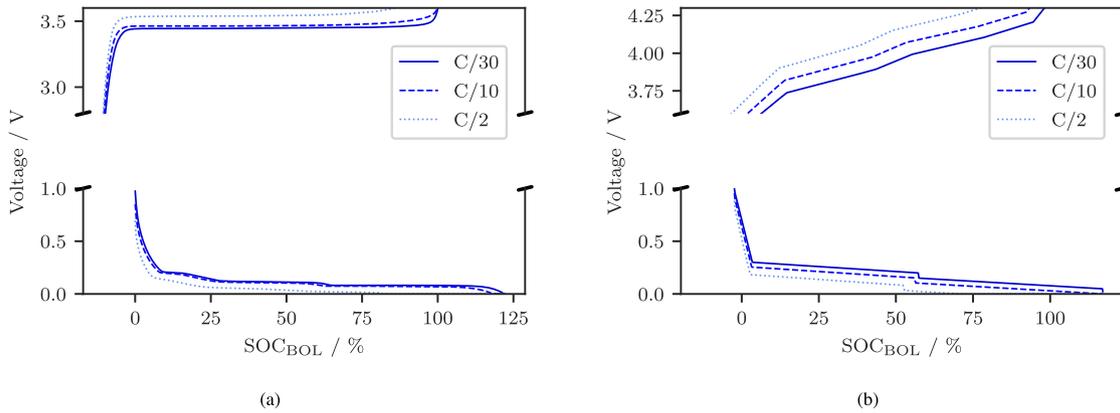
While this gives clear requirements for TL, it stays an open research question which features actually lead to accurate OCV reconstruction. Thus, we further elaborate on the hypothesis of gradient changes in the voltage trajectories as the main feature.

#### 4.2. Transfer learning from different cell chemistries

To validate our hypothesis that gradient changes is the sole requirement for transferring knowledge from the source to the target domain independent of its chemistry, we conduct additional experiments using two synthetic datasets. The synthetic datasets consist of constant current voltage curves, and Fig. 12 displays samples of the OCPs used to generate them. The samples include an LFP cathode with a graphite anode (Fig. 12(a)) and an artificial anode with an artificial cathode (Fig. 12(b)). It is important to note that the ART cell used in this study is completely self-designed and not the result of measurements. While the main trajectory is comparable to a conventional NMC- or NCA-graphite cell, the only information contained is a changing gradient at three and five positions in the anode and cathode OCP, respectively. Hence, this is the only transferable knowledge from the source domain and ultimately proves if this is the main feature for TL in the field of battery SOH estimation and OCV reconstruction.



**Fig. 11.** Evaluation of the TL-NN performance for UC3 and test data drawn from partial charging events at C/2 and 1C. For both cases, three degradation states and four different voltage/SOC windows are evaluated. The low SOC-range ((a)–(c)) corresponds to a voltage window from 3.25 V to 3.8 V, the middle window ((d)–(f)) from 3.6 V to 4.1 V and the high SOC-range ((g)–(i)) from 3.8 V to 4.2 V. Subfigures (j)–(l) show the results for the full input voltage. For all evaluated scenarios the mean absolute OCV reconstruction error is included in the subplot, in its respective color. The upper text belongs to the 1C event, while the lower text corresponds to the C/2 charge. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



**Fig. 12.** Additional synthetic dataset generated from half-cell potential curves at various C-rates. Two additional datasets are created, one for (a) LFP chemistry and another for (b) artificial (ART) chemistry.

In this study, we follow a consistent workflow, as depicted in Fig. 7, to investigate the effectiveness of plateau shifting in transferring knowledge between domains. The synthetic datasets used in our experiments are generated under identical conditions, including the same C-rates and aging paths. In contrast to our previous synthetic dataset, we utilized the alawa-toolbox from Dubarry et al. [6] to generate these constant current voltage curves. To ensure consistency and comparability across the datasets, we apply the same preprocessing steps, including voltage windowing, normalization, and data splitting. While the ART dataset uses the same defined voltage windows as

for the initial synthetic dataset in Table 2, the LFP cell undergoes a windowing with lower voltage levels: The minimum voltage  $V_{\min}$  is set to 2.8 V, 3.0 V, 3.2 V and 3.4 V. The maximum voltage  $V_{\max}$  varies between 3.4 V, 3.5 V and 3.6 V. To optimize the performance of the new base models, we employ Bayes tuning with the defined search space, as outlined in Table 3. This allows us to improve the models and achieve their respective optima. The final model architectures, along with their hyperparameters, are provided in the Appendix B to facilitate reproducibility and transparency. The base models are now trained and

**Table 5**

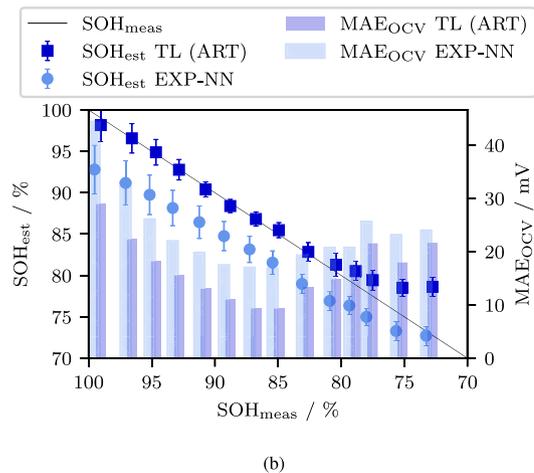
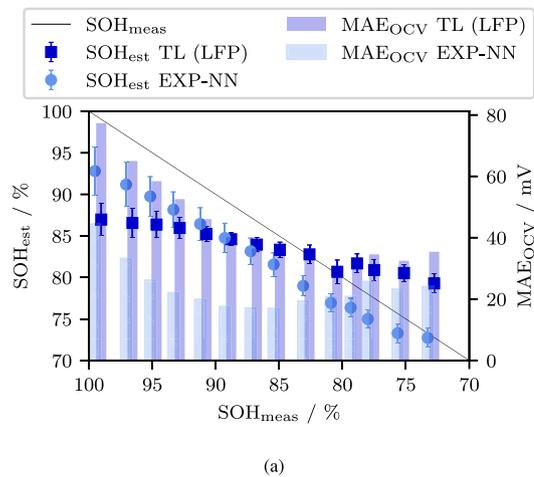
Performance comparison of the EXP-NN and the TL-NN (pretrained with a different cell chemistry) for different UCs in the voltage window between 3.6 V to 4.1 V. The UCs differ in the amount of training data for the fine-tuning process.

		MAPE <sub>SOH</sub>			MAE <sub>OCV</sub>		
		C/3.85	C/2	1C	C/3.85	C/2	1C
UC1	EXP-NN	-	2.0%	-	-	21 mV	-
	TL (LFP)	-	9.1%	-	-	50 mV	-
	TL (ART)	-	3.3%	-	-	21 mV	-
UC2	EXP-NN	2.8%	4.9%	2.5%	18 mV	18 mV	15 mV
	TL (LFP)	17.8%	8.6%	4.2%	78 mV	55 mV	26 mV
	TL (ART)	1.2%	5.4%	1.6%	11 mV	25 mV	11 mV
UC3	EXP-NN	7.2%	2.9%	3.3%	27 mV	22 mV	18 mV
	TL (LFP)	19.9%	15.3%	6.9%	74 mV	64 mV	40 mV
	TL (ART)	0.8%	4.7%	1.2%	12 mV	28 mV	11 mV
UC4	EXP-NN	1.2%	6.7%	20.4%	15 mV	140 mV	280 mV
	TL (LFP)	15.5%	8.3%	9.0%	66 mV	43 mV	43 mV
	TL (ART)	0.5%	5.6%	6.9%	11 mV	32 mV	47 mV

transfer-learned for the four defined UCs in Fig. 5. Specifically, we utilize fine-tuning with a reduced learning rate of  $1 \times 10^{-4}$  for 200 epochs. Noticeably, we must increase the learning rate in comparison to our initial TL-NN from  $1 \times 10^{-6}$  to  $1 \times 10^{-4}$  (Adam default value is  $1 \times 10^{-3}$ ) due to the decreased similarity of the source and target datasets. In the fine-tuning process, more adoptions of the network weights are mandatory to achieve sufficient learning of the target dataset which is not possible with a learning rate lower than  $1 \times 10^{-4}$ . The base models are compared to the EXP-NNs for each UC, and the final comparison results are summarized in Table 5. These results provide insights into the performance and effectiveness of plateau shifting in transferring knowledge across different UCs. The pretrained model from the LFP dataset fails to transfer its past knowledge to the target domain. For a voltage input window from 3.6 V to 4.1 V and UC1 to UC3 the TL LFP model performs worse than the benchmark EXP-NN. The LFP OCV curve is very flat and barely shows any detectable gradient changes in its trajectory. At higher C-rates the voltage charging curve seems even faded and does not provide any information to the model. The sole advantage is a moderate estimation for UC4. Nevertheless, the fine-tuning process does not allow a sufficient adaption to the target domain and hence the performance stays equally low for every UC. Fig. 13(a) illustrates the performance for the model on UC3 for the complete test dataset: The insufficient fine-tuning process is visible by the almost constant estimations over lifetime. The learning-rate or number of epochs, however, cannot be increased any higher because this will lead to catastrophic forgetting and completely erase the impact of the learned knowledge from the source dataset.

Contrary the ART model outperforms the reference EXP-NN, as can be seen in Fig. 13(b). The model achieves accurate estimations over lifetime with a small standard deviation. The OCV curve is mostly reconstructed with a MAE below 20 mV. The deviation, however, is remarkably higher than the initial TL pretrained from a similar cell chemistry (see Fig. 10). Table 5 provides further insights and reveals the advantage of the ART TL-NN for all use cases and the voltage window from 3.6 V to 4.1 V: While the model performs comparably to the EXP-NN for test data with C/2 and UC1 to UC3, it is very accurate for test data with C/3.85 and 1C. This adds on to our theory of including fixed boundaries of the target dataset to accelerate the transfer of knowledge. Although the test data for these C-rates includes different cells, inclusion of another cell at this specific C-rates for training drastically improves performance. This may be due to the better interpretation of C-rate dependent gradient changes and the location of those in the voltage curve.

In conclusion, the ART TL-NN exhibits slightly lower performance compared to the original TL-NN. This finding, however, emphasizes that the specific chemical composition of the source dataset is not a critical factor. Instead, it is the presence of gradients and its changes



**Fig. 13.** Evaluation of the EXP-NN and the TL-NNs pretrained with (a) LFP and (b) ART cell for UC3 and the whole test dataset. One data point corresponds to the average estimation within a 2% region and bins them for better comparability. The markers for the SOH estimations indicate the mean value, while the bars correspond to the standard deviation. Similar, the mean OCV reconstruction error is evaluated with the bar plot. To increase comparability between the EXP-NN and TL-NNs, the results from the TL ART model are shifted to the left by 0.25%, and the results from the EXP-NN are shifted to the right by 0.25%.

in the voltage curves that play a crucial role. Especially the cathode OCP heavily influences the performance because in the LFP dataset, the graphite anode shows some detectable gradient changes which are not sufficient to reconstruct the OCV accordingly. Consequently, the TCN-LSTM approach can effectively identify degradation modes and their underlying patterns in the voltage curves, enabling the mapping of these patterns to the altered OCV curve.

#### 4.3. Comprehensive discussion

The study at hand presents a TCN-LSTM network to reconstruct the full OCV curve from partial charging voltage segments at different C-rates. Our findings further confirm the research question that TL from synthetic data is possible and superior to traditional DL. For UC3 the TL-NN outperforms the more complex EXP-NN which are both trained / fine-tuned with 13.2% of the total experimental dataset. For every C-rate and a limited voltage window from 3.6 V to 4.1 V the TL-NN already reaches OCV-reconstruction MAEs below 20 mV which is in the range of literature values [4,11,13]. The SOH is calculated from the reconstructed OCV curves with MAPEs below 2.2%. We further highlight the influence of C-rate and voltage-range on the accuracy.

While lower voltage levels definitely lead to higher accuracies, the influence of the C-rate is obscured by the impact of the training data. Lower voltage segments may be beneficial because this allows an easier detection of the anode aging due to changing gradient of the voltage curve. Subsequent the identification of cathode aging is more straightforward due to the higher gradient in the OCP. By introducing two additional source synthetic datasets, we are able to answer the second research question about the actual learned knowledge from partial voltage curves: Our findings strongly suggest that the TCN-LSTM model comprehends the degradation patterns in the voltage curve, which are influenced by the shifting of plateaus caused by degradation modes.

In contrast to our findings, Bockrath et al. [16] concluded the middle and upper voltage-range to be more suitable for SOH estimation from partial discharge voltage segments. They claim that the middle and upper voltage ranges preserve more of the full capacity degradation measurements and that their learned knowledge differs from our hypothesis: Their TCN exploits the correlation between the measured capacity from partial measurements to the measured SOH. They show that their measured capacity throughput in the middle and upper voltage ranges correlates well with the SOH. This is not the main feature in our dataset, as we aim to reconstruct the full OCV curve and calculate the SOH from this curve. Hence, the sole information of the SOH is insufficient for accurate OCV reconstruction. Comparing our TL-NN from UC3 with the TCN developed by Bockrath et al. [16] for full voltage curves, we see that they have reached an SOH RMSE of 1.0%, while we were able to decrease the SOH MAPE to 0.64%. Zhou et al. [26] have for the first time introduced a TL approach for full OCV reconstruction from partial charging data. While they have successfully applied TL to different cell chemistries and reconstructed the OCV with errors below 2.73%, they failed to clarify the usability of synthetic data for the source domain. They have used a generative DL model and reached TL by fine-tuning the decoder basis. Contrary, we fine-tune our complete TCN-LSTM network by reducing the learning rate to  $1 \times 10^{-6}$  and epochs to a maximum of 200. Further, we have found no contributions towards a detailed analysis of the corresponding knowledge transferred between deep-learning models.

We prove that synthetic data indeed works as a source domain for TL, even for varying cells. Further, TL outperforms traditional DL as long as the fine-tuning data includes the boundaries of the target domain. TL increases its superiority against traditional DL with smaller training datasets. Generally, DL models for full OCV reconstruction reach higher accuracy for the lower voltage segments. The C-rate of the input data does not influence the performance as much as the training/testing split does. Hence, in most research studies, the clear influence of the C-rate is hard to identify. Pretraining from synthetic data sources and varying cell chemistries works as long as enough plateaus and gradient shifts are detectable in the voltage curves. Although pretraining the model with the source LFP dataset fails to detect any information, the use of an artificially designed dataset consisting of ART OCPs allows for the incorporation of knowledge regarding degradation modes and OCV degradation into the NN. Consequently, by including more detectable plateaus and a wider range of gradients in the voltage trajectory, the model is able to extract more information and exhibits a higher likelihood of successfully transferring its knowledge to a new target domain.

Table 4 shows that for UC2 and UC3 the performance is biased towards test data with C/3.85 and 1C. These data segments are included in the fine-tuning data and hence the possibility of overfitting arises on these parts of the data. This stands in contrast to the reduced learning rate and the smaller number of fine-tuning epochs. As Fig. 11 shows, the reconstructed OCV matches the true OCV well, even for test data with C/2 charging segments. Even though the MAE<sub>OCV</sub> for UC2 and UC3 are similar for the TL and the EXP-NN (see Table 4), the MAE<sub>SOH</sub> reveal a different result: The SOH estimation accuracy is higher for the TL-NN and hence proves a better fit of the whole reconstruction because

the MAE<sub>OCV</sub> is only calculated at  $Q$ -locations where both, the estimated and the true OCV, exist. Hence, a MAE<sub>OCV</sub> = 0 mV in combination with MAE<sub>SOH</sub>  $\geq$  10% is possible and thus the reconstructed curve may be incomplete.

Future research should tackle the exploration of more and other experimental datasets, including more C-rates, varying temperatures and more challenging cell chemistries, i.e., LFP or Si-containing anodes. The usage of synthetic data for TL arises the opportunity for DM estimation due to the contained labels in the synthetic dataset. While generating these labels in the experimental dataset is challenging, differential voltage analysis (DVA) and ICA with OCV measurements still can give rough insights about the degradation patterns. Similar, the output of the developed TL-NN can be fed into a mechanistic model approach [6] to reconstruct the full OCV curve based on shifting and scaling of the pristine OCPs. This further enables to estimate the DMs and includes more physico-chemical relevance into the estimation.

## 5. Conclusion

The study proposes a data-driven method to reconstruct the full OCV curve from partial charging voltage segments at different C-rates. We prove the effectiveness of TL from synthetic data and evaluate the limitations of this shortcoming by comparing our TL models for four different UCs to purely experimental ML models. In every UC the experimental training/fine-tuning data gets reduced from 67% for UC1 to 40% for UC2, to 13% for UC3 until only 6.6% of the data are included for the fourth and last UC4. While the training/fine-tuning data for UC1 to UC3 incorporate the boundaries of the dataset, i.e., highest and lowest SOH and C-rate, data is limited to the lowest C-rate for UC4. The TL-NN with a simpler architecture outperforms the EXP-NN for every UC and reaches a MAE<sub>OCV</sub> below 22 mV for the first three UCs and a partial voltage input curve from 3.6 V to 4.1 V. We conclude that a minimum voltage window spanning 400 mV is necessary to enable accurate OCV reconstruction. Further, the lower voltage segments are better suited for full OCV curve reconstruction due to the better recognition of features from the anode half-cell potential curve in the voltage segment. These segments allow accurate SOH estimation with errors below 14 mV MAE<sub>OCV</sub> and 1.1% MAPE<sub>SOH</sub>, respectively, for the first three UCs and  $V_{\min} \leq 3.4$  V. Reducing the training data amount to a minimum, requires the inclusion of the boundaries, i.e., the highest and lowest C-rate and SOH, in the fine-tuning datasets. We further find the influence of the C-rate to be negligible in our experiment but it may be a result of the train/test split. Pretraining from different source datasets and incorporating other cell chemistries allows for some interpretation of the actual learning patterns in the black-box TCN-LSTM. Using a completely artificial dataset with many plateaus and gradient switches allows the model to comprehend the theory of DMs and apply this prelearned knowledge to the new target domain. Pretraining from the LFP dataset, however, fails due to no detectable gradient changes in the voltage trajectory.

The positive results of the TL-NN suggest the general usage of TL from synthetic data for battery state estimation. Especially for SOC and SOH estimation over lifetime, the generation of comprehensive data at various aging paths might accelerate the industrialization for BEV fleets and the low-cost development of data-driven state estimation algorithms by reducing measurement time and cost. The generation of synthetic data for various batteries is publicly available (see alawa-toolbox [6]) and allows the pretraining of ML models for multiple UCs. Even the deployment of ML models for new cell chemistries is feasible due to the low requirements of fine-tuning data from the target domain. Hence, this research project takes a promising step towards generalizable ML models for almost every battery type.

The limitations of TL, however, must be kept in mind by fulfilling the minimum requirements of the pretraining and fine-tuning datasets. The application of ML models for OCV curve reconstruction from

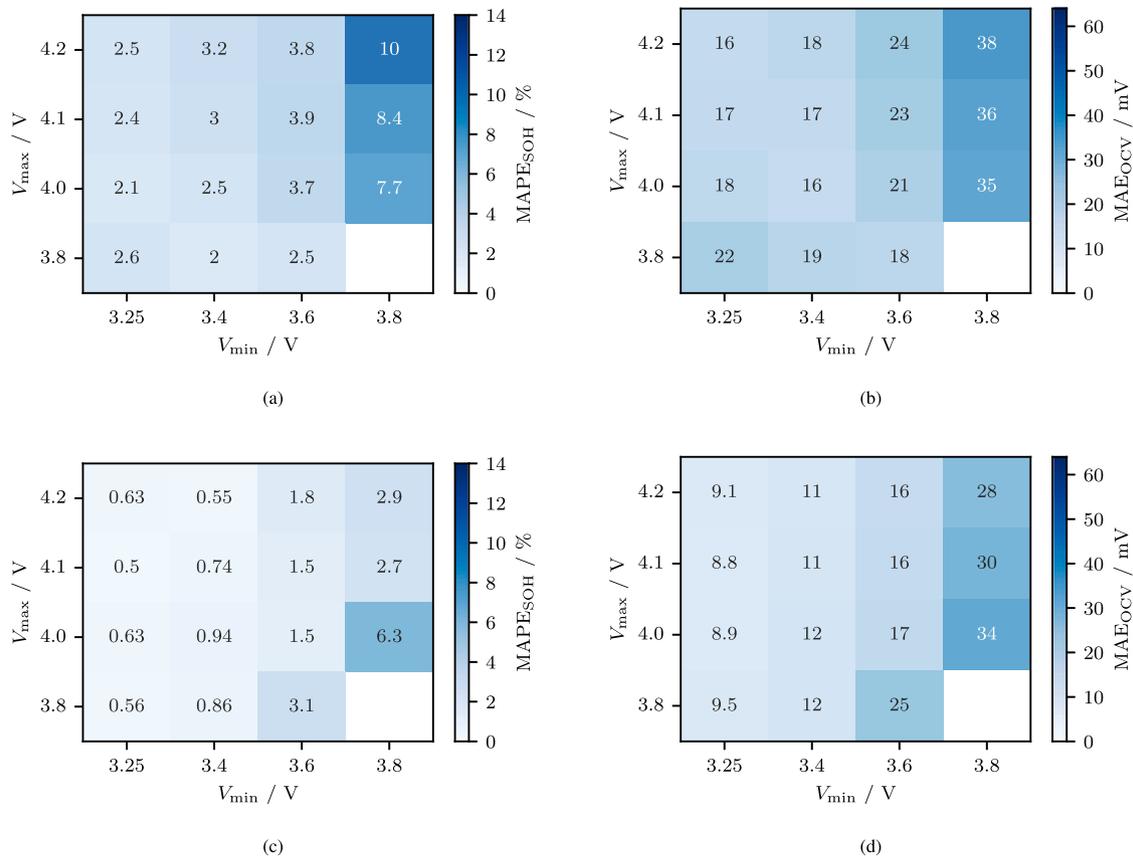


Fig. A.14. Results for UC2: (a), (b) EXP-NN, (c), (d) TL-NN. (a) Average SOH error for the test dataset with the reference EXP-NN. (b) Average OCV error for the test dataset with the reference EXP-NN. (c) Average SOH error for the test dataset with the TL-NN. (d) Average OCV error for the test dataset with the TL-NN.

charging voltage segments is generally limited to batteries with sufficient voltage trajectories, i.e., huge challenges arise for LFP cells. The application of TL to other cell chemistries in the target domain remains an open research question and must be explored in future studies. Hence, the inclusion of larger and more diverse experimental datasets in the target domain will answer many of the open questions. Moreover, the application of the method to real-world BEV data must be answered in the near future. Our proof-of-concept for TL-based SOH estimation opens new research perspectives for SOC, SOH and DM estimation. With the usage of synthetic data, many other state estimation models can be developed faster and more efficiently.

#### CRedit authorship contribution statement

**Tobias Hofmann:** Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Data curation, Conceptualization. **Jacob Hamar:** Writing – review & editing, Supervision. **Bastian Mager:** Investigation. **Simon Erhard:** Writing – review & editing, Supervision. **Jan Philipp Schmidt:** Writing – review & editing, Supervision.

#### Declaration of competing interest

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

#### Data availability

Data will be made available on request.

#### Acknowledgments

Funded by the Open Access Publishing Fund of the University of Bayreuth. This work was performed in cooperation with the University of Bayreuth - Chair of Systems Engineering for Electrical Energy Storage and BMW Group. The authors would like to thank the team of Batemo GmbH for providing the Batemo Cell Models as high-precision physical battery cell models to generate the simulation data. The authors would also like to thank Matthieu Dubarry and the University of Hawai'i at Mānoa for sharing the alawa battery emulation toolbox.

#### Appendix A. Detailed benchmark results

This section provides a more detailed analysis of each TL and EXP-NN for UC2 to UC4. The Figures provide insights about the OCV curve reconstruction and the SOH estimation accuracy with different partial voltage windows as input data.

##### A.1. Use Case 2

See Fig. A.14.

##### A.2. Use Case 3

See Fig. A.15.

##### A.3. Use Case 4

See Fig. A.16.

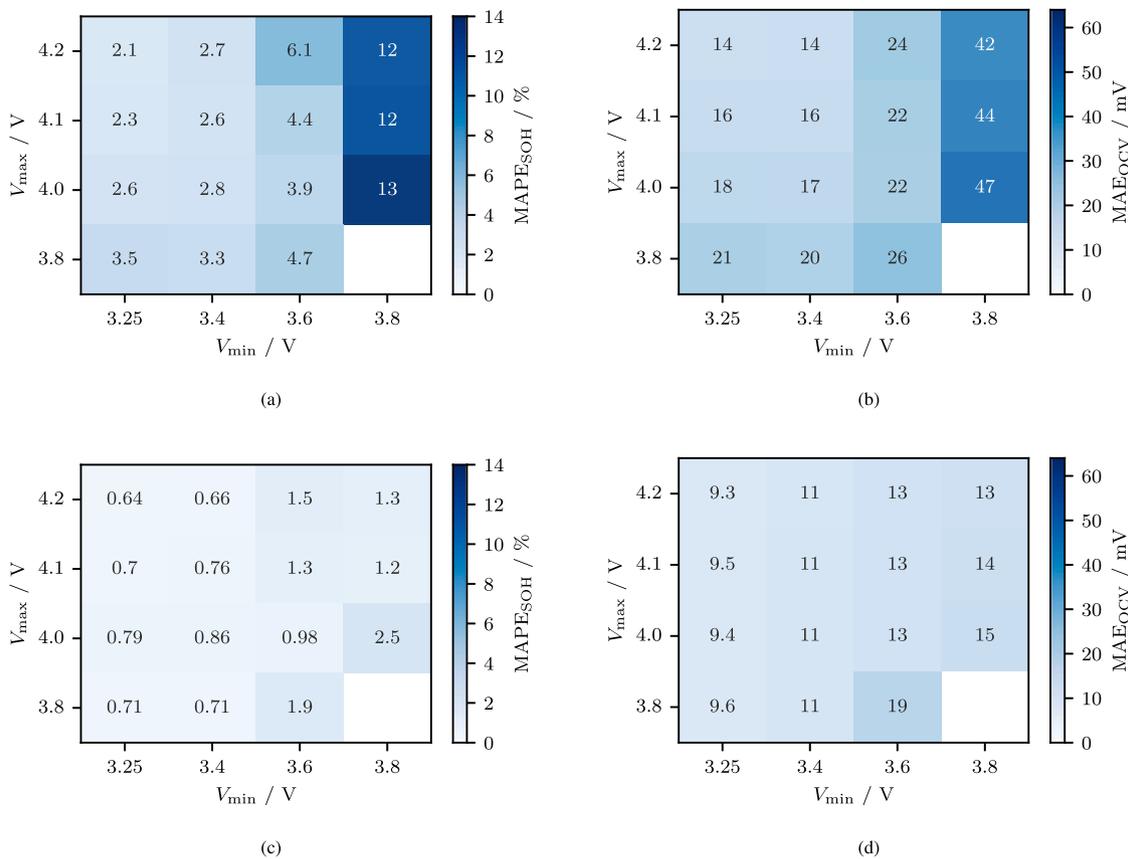


Fig. A.15. Results for UC3: (a), (b) EXP-NN, (c), (d) TL-NN. (a) Average SOH error for the test dataset with the reference EXP-NN. (b) Average OCV error for the test dataset with the reference EXP-NN. (c) Average SOH error for the test dataset with the TL-NN. (d) Average OCV error for the test dataset with the TL-NN.

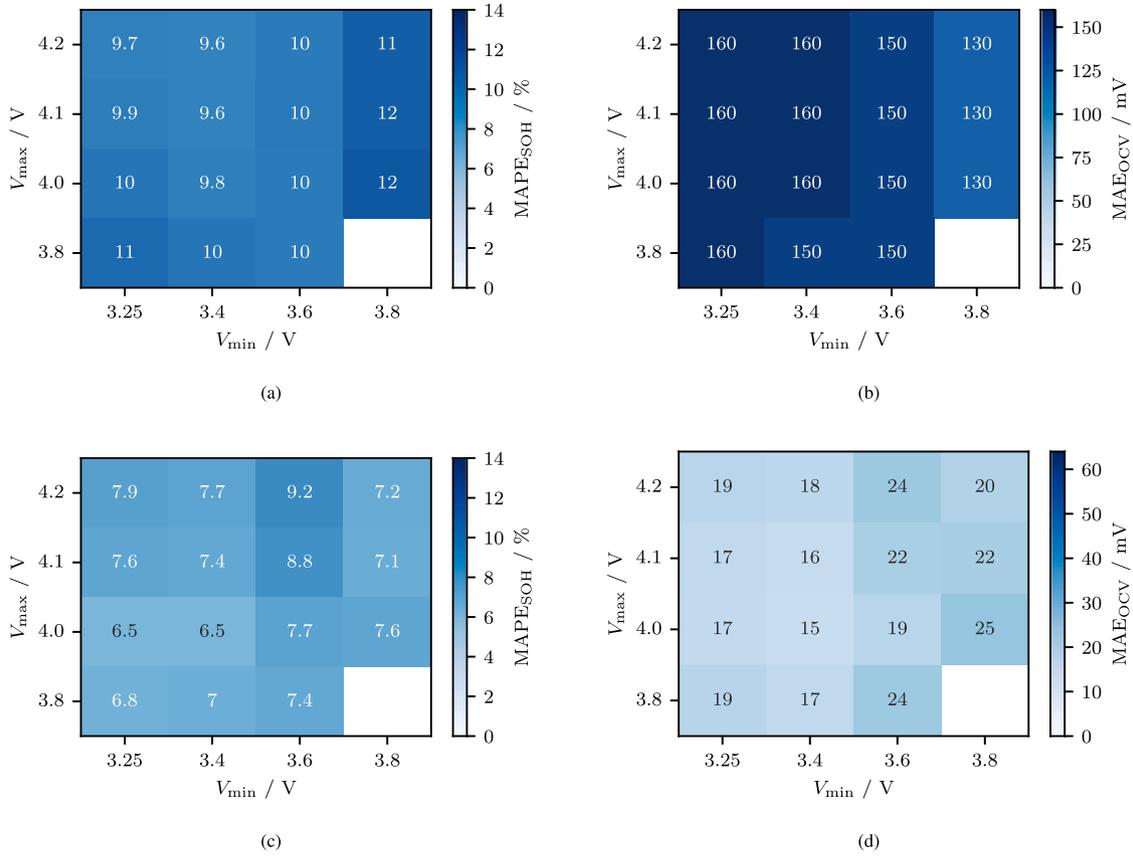


Fig. A.16. Results for UC4: (a), (b) EXP-NN, (c), (d) TL-NN. (a) Average SOH error for the test dataset with the reference EXP-NN. (b) Average OCV error for the test dataset with the reference EXP-NN. (c) Average SOH error for the test dataset with the TL-NN. (d) Average OCV error for the test dataset with the TL-NN.

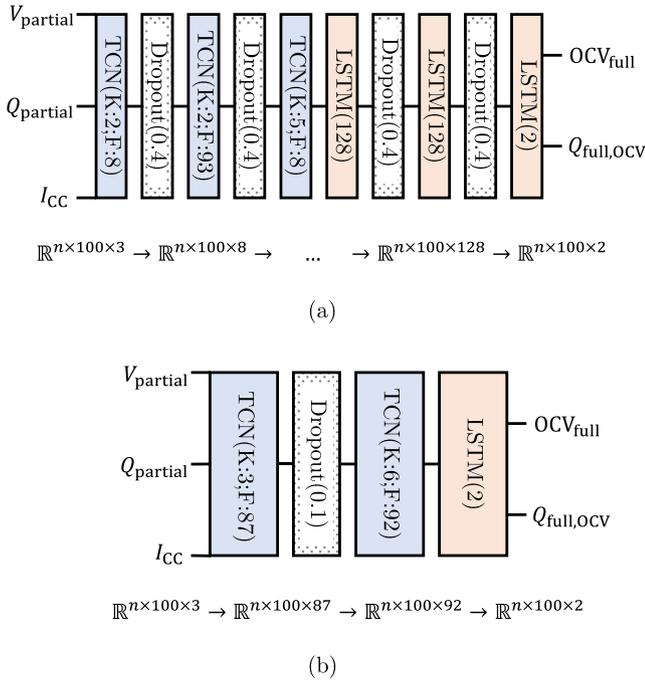


Fig. B.17. Final architecture of the tuned synthetic models optimized with data from the (a) LFP dataset and the (b) ART dataset and the corresponding dimensions of output data from each layer with  $n$  as the number of samples. In the TCN-layers the first number refers to the kernel size and the second number to the filter size.

## Appendix B. Model architecture of the LFP and ART model

Fig. B.17 provides insights about the final model architectures trained from the additional source datasets LFP and ART. The final architectures are a result of an extensive Bayesian hyperparameter optimization with 100 trials.

## References

- [1] Sauer DU, Bopp G, Jossen A, Garche J, Rothert M, Wollny M. State of Charge — What do we really speak about. In: The 21st international telecommunications energy conference. 1999, p. 6–9.
- [2] Plett GL. Extended Kalman filtering for battery management systems of LiPB-based HEV battery packs: Part 2. Modeling and identification. J Power Sources 2004;134(2):262–76. <http://dx.doi.org/10.1016/j.jpowsour.2004.02.032>.
- [3] Plett GL. Extended Kalman filtering for battery management systems of LiPB-based HEV battery packs: Part 3. State and parameter estimation. J Power Sources 2004;134(2):277–92. <http://dx.doi.org/10.1016/j.jpowsour.2004.02.033>.
- [4] Schmitt J, Rehm M, Karger A, Jossen A. Capacity and degradation mode estimation for lithium-ion batteries based on partial charging curves at different current rates. J Energy Storage 2023;59:106517. <http://dx.doi.org/10.1016/j.est.2022.106517>.
- [5] Hofmann T, Li J, Hamar J, Erhard S, Schmidt JP. The  $4Q$ -method: State of health and degradation mode estimation for lithium-ion batteries using a mechanistic model with relaxed voltage points. J Power Sources 2024;596:234107. <http://dx.doi.org/10.1016/j.jpowsour.2024.234107>.
- [6] Dubarry M, Truchot C, Liaw BY. Synthesize battery degradation modes via a diagnostic and prognostic model. J Power Sources 2012;219:204–16. <http://dx.doi.org/10.1016/j.jpowsour.2012.07.016>.
- [7] Zhang C, Zhao S, Yang Z, Chen Y. A reliable data-driven state-of-health estimation model for lithium-ion batteries in electric vehicles. Front Energy Res 2022;10. <http://dx.doi.org/10.3389/fenrg.2022.1013800>.
- [8] Zhang C, Luo L, Yang Z, Zhao S, He Y, Wang X, Wang H. Battery SOH estimation method based on gradual decreasing current, double correlation analysis and GRU. Green Energy Intell Transp 2023;2(5):100108. <http://dx.doi.org/10.1016/j.geits.2023.100108>.

- [9] Zhang C, Luo L, Yang Z, Du B, Zhou Z, Wu J, Chen L. Flexible method for estimating the state of health of lithium-ion batteries using partial charging segments. *Energy* 2024;295:131009. <http://dx.doi.org/10.1016/j.energy.2024.131009>.
- [10] Dubarry M, Costa N, Matthews D. Data-driven direct diagnosis of Li-ion batteries connected to photovoltaics. *Nature Commun* 2023;14:3138. <http://dx.doi.org/10.1038/s41467-023-38895-7>.
- [11] Tian J, Xiong R, Shen W, Sun F. Electrode ageing estimation and open circuit voltage reconstruction for lithium ion batteries. *Energy Storage Mater* 2021;37:283–95. <http://dx.doi.org/10.1016/j.ensm.2021.02.018>.
- [12] Ruan H, Chen J, Ai W, Wu B. Generalised diagnostic framework for rapid battery degradation quantification with deep learning. *Energy AI* 2022;9:100158. <http://dx.doi.org/10.1016/j.eyai.2022.100158>.
- [13] Guo R, Xu Y, Hu C, Shen W. A curve relocation approach for robust battery open circuit voltage reconstruction and capacity estimation based on partial charging data. *IEEE Trans Power Electron* 2024;39(3):3760–73. <http://dx.doi.org/10.1109/TPEL.2023.3347236>.
- [14] O'Shea K, Nash R. An introduction to convolutional neural networks. 2015. <http://dx.doi.org/10.48550/arXiv.1511.08458>, [arXiv:1511.08458](https://arxiv.org/abs/1511.08458).
- [15] Lea C, Flynn MD, Vidal R, Reiter A, Hager GD. Temporal convolutional networks for action segmentation and detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition. CVPR, 2017, p. 156–65. <http://dx.doi.org/10.48550/arXiv.1611.05267>.
- [16] Bockrath S, Lorentz V, Pruckner M. State of health estimation of lithium-ion batteries with a temporal convolutional neural network using partial load profiles. *Appl Energy* 2023;329:120307. <http://dx.doi.org/10.1016/j.apenergy.2022.120307>.
- [17] Li C, Han X, Zhang Q, Li M, Rao Z, Liao W, Liu X, Liu X, Li G. State-of-health and remaining-useful-life estimations of lithium-ion battery based on temporal convolutional network-long short-term memory. *J Energy Storage* 2023;74(Part B):109498. <http://dx.doi.org/10.1016/j.est.2023.109498>.
- [18] Liu S, Chen Z, Yuan L, Xu Z, Jin L, Zhang C. State of health estimation of lithium-ion batteries based on multi-feature extraction and temporal convolutional network. *J Energy Storage* 2024;75:109658. <http://dx.doi.org/10.1016/j.est.2023.109658>.
- [19] Zhou D, Li Z, Zhu J, Zhang H, Hou L. State of health monitoring and remaining useful life prediction of lithium-ion batteries based on temporal convolutional network. *IEEE Access* 2020;8:53307–20. <http://dx.doi.org/10.1109/ACCESS.2020.2981261>.
- [20] Zhou D, Wang B. Battery health prognosis using improved temporal convolutional network modeling. *J Energy Storage* 2022;51:104480. <http://dx.doi.org/10.1016/j.est.2022.104480>.
- [21] Tan C, Sun F, Kong T, Zhang W, Yang C, Liu C. A survey on deep transfer learning. In: Artificial neural networks and machine learning – ICANN. 2018, p. 11141. [http://dx.doi.org/10.1007/978-3-030-01424-7\\_27](http://dx.doi.org/10.1007/978-3-030-01424-7_27).
- [22] Sahoo S, Hariharan KS, Agarwal S, Swernath SB, Bharti R, Han S, Lee S. Transfer learning based generalized framework for state of health estimation of Li-ion cells. *Sci Rep* 2022;12:13173. <http://dx.doi.org/10.1038/s41598-022-16692-4>.
- [23] Zou G, Yan Z, Zhang C, Song L. Transfer learning with CNN-LSTM model for capacity prediction of lithium-ion batteries under small sample. *J Phys Conf Ser* 2022;2258:012042. <http://dx.doi.org/10.1088/1742-6596/2258/1/012042>.
- [24] Shen S, Sadoughi M, Li M, Wang Z, Hu C. Deep convolutional neural networks with ensemble learning and transfer learning for capacity estimation of lithium-ion batteries. *Appl Energy* 2020;260:114296. <http://dx.doi.org/10.1016/j.apenergy.2019.114296>.
- [25] Tian J, Ma L, Zhang T, Han T, Mai W, Chung CY. Exploiting domain knowledge to reduce data requirements for battery health monitoring. *Energy Storage Mater* 2024;67:103270. <http://dx.doi.org/10.1016/j.ensm.2024.103270>.
- [26] Zhou Z, Liu Y, Zhang C, Shen W, Xiong R. Deep neural network-enabled battery open-circuit voltage estimation based on partial charging data. *J Energy Chem* 2024;90:120–32. <http://dx.doi.org/10.1016/j.jechem.2023.11.009>.
- [27] Keil P, Schuster SF, Wilhelm J, Travi J, Hauser A, Karl RC, Jossen A. Calendar aging of lithium-ion batteries. *J Electrochem Soc* 2016;163(9):A1872. <http://dx.doi.org/10.1149/2.0411609jes>.
- [28] Hamar JC, Stuckenberger M, Sturm J, Schmitt J, Rogge M, Erhard SV, Jossen A. Investigating the path dependent aging behavior of nickel cobalt aluminum oxide cathode batteries during high C-rate cycling conditions. *J Electrochem Soc* 2024;171(2):020555. <http://dx.doi.org/10.1149/1945-7111/ad2952>.
- [29] Goodfellow I, Bengio Y, Courville A. Deep learning. MIT Press; 2016. <http://www.deeplearningbook.org>.
- [30] Hamar JC, Erhard SV, Canesso A, Kohlschmidt J, Olivain N, Jossen A. State-of-health estimation using a neural network trained on vehicle data. *J Power Sources* 2021;512:230493. <http://dx.doi.org/10.1016/j.jpowsour.2021.230493>.
- [31] Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput* 1997;9(8):1735–80. <http://dx.doi.org/10.1162/neco.1997.9.8.1735>.
- [32] Li W, Sengupta N, Dechent P, Howey D, Annaswamy A, Sauer DU. Online capacity estimation of lithium-ion batteries with deep long short-term memory networks. *J Power Sources* 2021;482:228863. <http://dx.doi.org/10.1016/j.jpowsour.2020.228863>.
- [33] Gong Q, Wang P, Cheng Z. A data-driven model framework based on deep learning for estimating the states of lithium-ion batteries. *J Electrochem Soc* 2022;169:030532. <http://dx.doi.org/10.1149/1945-7111/ac5bac>.
- [34] Bai S, Koller JZ, Koltun V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. 2018. <http://dx.doi.org/10.48550/arXiv.1803.01271>, [arXiv preprint](https://arxiv.org/abs/1803.01271).
- [35] Chemali E, Kollmeyer PJ, Preindl M, Fahmy Y, Emadi A. A convolutional neural network approach for estimation of li-ion battery state of health from charge profiles. *Energies* 2022;15(3):1185. <http://dx.doi.org/10.3390/en15031185>.
- [36] Qian C, Xu B, Chang L, Sun B, Feng Q, Yang D, Ren Y, Wang Z. Convolutional neural network based capacity estimation using random segments of the charging curves for lithium-ion batteries. *Energy* 2021;227:120333. <http://dx.doi.org/10.1016/j.energy.2021.120333>.
- [37] Xu H, Wu L, Xiong S, Li W, Garg A, Gao L. An improved CNN-LSTM model-based state-of-health estimation approach for lithium-ion batteries. *Energy* 2023;276:127585. <http://dx.doi.org/10.1016/j.energy.2023.127585>.
- [38] Weber M, Auch M, Doblender C, Mandl P, Jacobsen H-A. Transfer learning with time series data: A systematic mapping study. *IEEE Access* 2021;9:165409–32.
- [39] Lu Z, Yu XL, Wei LC, Cao F, Zhang LY, Meng XZ, Jin LW. A comprehensive experimental study on temperature-dependent performance of lithium-ion battery. *Appl Therm Eng* 2019;158:113800. <http://dx.doi.org/10.1016/j.applthermaleng.2019.113800>.
- [40] Sola J, Sevilla J. Importance of input data normalization for the application of neural networks to complex industrial problems. *IEEE Trans Nucl Sci* 1997;44(3):1464–8. <http://dx.doi.org/10.1109/23.589532>.
- [41] Victoria AH, Maragatham G. Automatic tuning of hyperparameters using Bayesian optimization. *Evol Syst* 2021;12:217–23. <http://dx.doi.org/10.1007/s12530-020-09345-2>.
- [42] O'Malley T, Bursztein E, Long J, Chollet F, Jin H, Invernizzi L, et al. Keras tuner. GitHub; 2019. <https://github.com/keras-team/keras-tuner>.
- [43] Schindler S, Baure G, Danzer MA, Dubarry M. Kinetics accommodation in Li-ion mechanistic modeling. *J Power Sources* 2019;440:227117. <http://dx.doi.org/10.1016/j.jpowsour.2019.227117>.
- [44] Yang S, Zhang C, Jiang J, Zhang W, Gao Y, Zhang L. A voltage reconstruction model based on partial charging curve for state-of-health estimation of lithium-ion batteries. *J Energy Storage* 2021;35:102271. <http://dx.doi.org/10.1016/j.est.2021.102271>.
- [45] Chen J, Marlow MN, Jiang Q, Wu B. Peak-tracking method to quantify degradation modes in lithium-ion batteries via differential voltage and incremental capacity. *J Energy Storage* 2022;45:103669. <http://dx.doi.org/10.1016/j.est.2021.103669>.
- [46] Lu D, Trimboli MS, Fan G, Zhang R, Plett GL. Implementation of a physics-based model for half-cell open-circuit potential and full-cell open-circuit voltage estimates: Part II. Processing full-cell data. *J Electrochem Soc* 2021;168:070533. <http://dx.doi.org/10.1149/1945-7111/ac11a5>.
- [47] Cui Z, Cui N, Li C, Lu J, Zhang C. Online identification and reconstruction of open-circuit voltage for capacity and electrode aging estimation of lithium-ion batteries. *IEEE Trans Ind Electron* 2023;70:4716–26. <http://dx.doi.org/10.1109/TIE.2022.3187596>.