State of Health Forecasting of Heterogeneous Lithium-ion Battery Types and Operation Enabled by Transfer Learning

Friedrich von Bülow[1], and Tobias Meisen[2]

[1]*Volkswagen AG, Berliner Ring 2, 38436 Wolfsburg, Germany*
friedrich.von.buelow@volkswagen.de

[2]*Institute of Technologies and Management of the Digital Transformation, University of Wuppertal, Rainer-Gruenter-Str. 21, 42119, Wuppertal, Germany*
meisen@uni-wuppertal.de

ABSTRACT

Due to the global transition to electromobility and the associated increased use of high-performance batteries, research is increasingly focused on estimating and forecasting the state of health (SOH) of lithium-ion batteries. Several data-intensive and well-performing methods for SOH forecasting have been introduced. However, these approaches are only reliable for new battery types, e.g., with a new cell chemistry, if a sufficient amount of training data is given, which is rarely the case. A promising approach is to transfer an established model of another battery type to the new battery type, using only a small amount of data of the new battery type. Such methods in machine learning are known as transfer learning. The usefulness and applicability of transfer learning and its underlying methods have been very successfully demonstrated in various fields, such as computer vision and natural language processing. Heterogeneity in battery systems, such as differences in rated capacity, cell cathode materials, as well as applied stress from use, necessitate transfer learning concepts for data-based battery SOH forecasting models. Hereby, the general electrochemical behavior of lithium-ion batteries, as a major common characteristic, supposedly provides an excellent starting point for a transfer learning approach for SOH forecasting models. In this paper, we present a transfer learning approach for SOH forecasting models using a multilayer perceptron (MLP). We apply and evaluate it on the method presented by von Bülow, Mentz, and Meisen (2021) using five battery datasets. In this regard, we investigate the optimal conditions and settings for the development of transfer learning with respect to suitable data from the target domain, as well as hyperparameters such as learning rate and frozen layers. We show that for the transfer of a SOH forecasting model to a new battery type it is more beneficial to have data of few old batteries, compared to data of many new batteries, especially in the case of superlinear degradation with knee points. Contrarily to computer vision freezing no layers is preferable in 95% of the experimental scenarios.

Friedrich von Bülow et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 United States License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

1. INTRODUCTION

Due to the mobility transition to electric vehicles worldwide, the battery's state of health (SOH) gains interest of customers and consequently by research and industry. The SOH reflects the battery ageing which depends on the usage and environmental conditions of the battery. SOH forecasting enables e.g. fleet managers of battery electric vehicle (BEV) fleets to optimize their operational strategy w.r.t. battery ageing. In addition, by forecasting the replacement time of old BEVs due to battery degradation the transition to a new vehicle type can be supported. Nevertheless, establishing such services, requires reliability and availability in the underlying forecasting models. For example, when launching new BEVs with a new battery type, a usable model satisfying the aforementioned requirements is required.

In this regard, both von Bülow et al. (2021) and Richardson, Osborne, and Howey (2019) have published reliable models for SOH prediction, though both are data-intensive. However, this is not justifiable for application in changing environments in most cases, as new data to build a reliable model using these methods cannot be collected in a temporal manner. Hence, due to the aforementioned needed availability their wide acceptance is currently limited. This is especially problematic for new battery types whose data availability is limited in the initial phase: First, only few laboratory battery ageing test are conducted which suffer under limited comparability to real-world operational conditions (Nuhic, Terzimehic, Soczka-Guth, Buchholz, and Dietmayer 2013; Sulzer et al. 2021; von Bülow and Meisen 2022). Second, usually battery cells, but not packs, systems, or modules are tested in the laboratory. Third, potentially only a few prototypes or endurance tests may have been operated using the new cell type.

One obvious solution to extend this limited data base, is further extensive data generation either in the laboratory or using endurance test vehicles. Data synthesis or augmentation from physical ageing models still needs validation from laboratory tests. Furthermore, we consider an adaption of feature values of ageing data and knowing the resulting ageing as too difficult. The complexity of this causal relationship is the main reason why data-driven models are considered for SOH forecasting. Nevertheless, these solutions are expensive and difficult as battery ageing is a lengthy process. In their work, von Bülow et al. (2021) proposed a solution by transferring an established model for battery ageing of another battery type to the new battery type, as soon as a small amount of data of the new battery type has been gathered. Such methods, summarized under the term transfer learning, have been successfully applied in different domains, like computer vision (Shao, Zhu, and Li 2015) and natural language processing (NLP) (Ruder, Peters, Swayamdipta, and Wolf 2019). The application of transfer learning for battery SOH forecasting models is a crucial part, as there are differences in batteries like the nominal capacity, the cell cathode materials as well as the applied load due to usage. However, the general electrochemical behavior of lithium-ion batteries is a major common characteristic which supposedly provides an excellent starting point for transfer learning. As von Bülow and Meisen (2022) state, until now transfer learning has not yet been applied to battery SOH forecasting.

In this regard, we contribute a transfer learning training process for SOH forecasting using a multilayer perceptron (MLP) with comprehensive experiments on five different known public datasets. We use the model presented by von Bülow et al. (2021) which showed very good results without limited availability of data. In the following, we examine when and how to transfer for. “When to transfer” concerns data availability in the target domain. I.e. the number of samples and their distribution by quantity and age of batteries. This shall enable practitioners to estimate the amount of data required for a successful transfer. “How to transfer” concerns the transfer method (i.e. freezing of layers of the pre-trained source model).

The remainder of this paper is structured as follows: Section 2 introduces the state of the art of battery components, their ageing and transfer learning including its applications in computer vison and battery ageing. In Section 3, the methods for SOH forecasting and transfer learning are explained. The used data basis is presented in Section 4. Subsequent, we present and discuss our results in Section 5. Section 6 concludes our work.

2. State of the Art

2.1. Lithium-Ion Battery Components and Ageing

The major components of a lithium-ion battery cell are: A negative electrode (anode), a positive electrode (cathode), the ion-conducting electrolyte, and the electrically insulating separator. For a schematic representation of a typical lithium-ion battery cell and information on the operating principle, interested readers are referred to (Keil 2017; Leuthner 2018; von Srbik 2015). The traditional cathode material has been lithium cobalt oxide (LCO). Alternatives are lithium nickel manganese cobalt oxide (NMC) and lithium iron phosphate (LFP) which have advantages over LCO regarding safety, cost, and size. Nevertheless regarding ageing, lithium-ion battery cells with different materials have different ageing characteristics (Vuorilehto 2018). Thereby, battery ageing is usually measured by the degradation of SOH. The SOH can be described by the internal resistance (SOH_{R}) and the remaining capacity (SOH_{C}) (Chen, Lü, Lin, Li, and Pan 2018; Waag, Fleischer, and Sauer 2014). The relative change of internal ohmic resistance compared to a new battery is the SOH_{R}. The capacity-based SOH_C is the remaining capacity $C(t)$ relative to the initial capacity of a new battery, also called nominal capacity C_{nom} (Lipu et al. 2018):

$$SOH(t) = SOH_{\mathrm{C}}(t) = \frac{C(t)}{C_{\mathrm{nom}}}. \tag{1}$$

In the following, we focus on the SOH_{C}, for simplicity referred to as SOH.

Battery ageing can be structured into two causes considered in this paper: Calendar ageing and cyclic ageing. Calendar ageing is associated with the storage of batteries, meaning no charging or discharging is applied. Hence, it is also called passive ageing. Cyclic aging corresponds to the impact of battery usage on the SOH, i.e. ageing due to charging and discharging (Gewald et al. 2020).

High temperatures (T) and high state of charges (SOC) are causing fast battery calendar and cyclic ageing (Matadi et al. 2017). For example, a high SOC over 80% accelerates solid electrolyte interphase (SEI) growth (Barré et al. 2013). Other stressors accelerating battery ageing are high charge and discharge C-rates[1] as well as a high ΔSOC (Gewald et al. 2020; Marongiu, Roscher, and Sauer 2015). Known battery stressors are qualitatively displayed in Table 1.

For the datasets used in this work, two types of battery ageing trajectories are relevant: Linear and superlinear degradation. Linear degradation is characterized by a constant aging rate over the whole battery life (e.g. NASA random in Appendix Figure 9). In contrast, batteries with superlinear degradation first age slowly, but change to accelerated ageing after the

[1] C-rate in [1/h]=[A/Ah] is the current relative to the nominal capacity C_{nom}.

Table 1. Aging mechanisms and their accelerating stressors (Birkl 2017; Nguyen 2019)

Battery component	Aging mechanisms	Accelerated by
Anode	Lithium plating	↑ C-rate, ↓ T, ↑ SOC
	Electrolyte decomposition	↑ T, ↑ SOC
	SEI formation	↑ & ↓ SOC
	SEI decomposition	↑ C-rate, ↓ T
	SEI growth	↑ T, ↑ SOC
	Structural disordering	↑ C-rate, ↑ & ↓ SOC
	Corrosion and loss of electrical contact	↓ SOC
Separator	Blocked pores (Separator and Electrodes)	↑ T, ↑ SOC
Cathode	Dissolving of transition metals	↑ T,
	Binder decomposition	↑ T, ↑ SOC
	Structural disordering	↑ C-rate, ↑ T
	Corrosion and loss of electrical contact	↑ C-rate, ↑ T, ↑ SOC

knee-point (e.g. Data-Driven in Appendix Figure 9) (Attia et al. 2022; Fermín-Cueto et al. 2020).

2.2. Transfer Learning

Transfer learning is a learning paradigm in machine learning that utilizes knowledge previously attained in one domain to solve a task in a novel domain (Pan and Yang 2010).

A domain $D = \{\mathcal{X}, P(X)\}$ is defined by a feature space $\mathcal{X}$ and a marginal probability distribution $P(X)$, where $X = \{x_1, \dots, x_n\}, x_i \in \mathcal{X}$. X represents a particular sample that is made up of different observations x_i which all lie in the feature space $\mathcal{X}$.

A task $T = \{\Upsilon, f(\cdot)\}$ is defined by a label space Υ and an objective predictive function $f(\cdot)$, which is learned from the training data. $f(\cdot)$ can be seen as $P = (y|x)$ from a probabilistic perspective.

In general, we assume that there is source domain data as $D_S = \{(x_{S_1}, y_{S_1}), \dots, (x_{S_n}, y_{S_n})\}$, where $x_{S_i} \in \mathcal{X}_S$ is an input vector and $y_{S_i} \in \Upsilon_S$ is the corresponding output vector. The target domain data can be denoted as $D_T = \{(x_{T_1}, y_{T_1}), \dots, (x_{T_n}, y_{T_n})\}$, where $x_{T_i} \in \mathcal{X}_T$ is an input vector and $y_{T_i} \in \Upsilon_T$ is the corresponding output vector. (Pan and Yang 2010)

Using these formal definitions, transfer learning can be defined in the follow way:

"Given a source domain D_S and learning task T_s, target domain D_T and learning task T_T, transfer learning aims to help improve the learning of the target predictive function $f(\cdot)$ in D_T using the knowledge in D_S and T_s, where $D_S \neq D_T$, or $T_S \neq T_T$." (Pan and Yang 2010)

When applying transfer learning, Torrey and Shavlik (2010) indentify three measures by which learning might be improved due to a transfer. These are visualized in Figure 1 at the exemplary model performance of a regression task measured by the root mean squared error (RMSE): Start, convergence speed and the asymptote of the model's performance may improve. First, the initial RMSE before any training with data from the target domain might be lower with transfer learning than without. This occurs when a model, i.e. an artificial neural network (ANN) trained on source domain data provides a better starting for learning than a randomly initialized model. Second, transfer learning can increase the convergence speed, i.e. it can decrease the training time to fully learn the target task. Third, the final performance achievable in the target task with transfer learning can be lower than without transfer learning. The lower start and lower asymptote can easily be measured by the model performance at the start and end of training respectively. However, measuring the higher convergence speed is more difficult. A popular metric is the area under the learning curve (AULC). For the convergence speed we are interested in the improvement of the model performance. Thus, the exponent

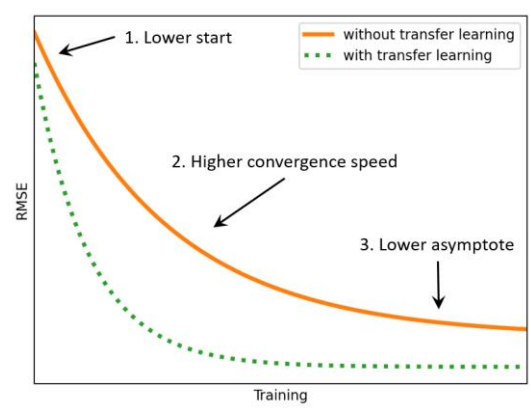


Figure 1: Measures in which transfer might improve learning: 1. lower start, 2. higher convergence speed and 3. lower asymptote (adapted from Torrey and Shavlik 2010).

of a power-law fit is more suitable (Bertoldi, Cettolo, Federico, and Buck 2012; Viering and Loog 2021). Mathematically, this measure is concerning relative convexity (Palmer 2003). However, overall we consider the lower asymptote as the most important measure because the final model performance after transfer learning is the most relevant. Thus, we focus our experiments and their discussion on the final model performance.

Further, Pan and Yang (2010) name three main concerns relevant when implementing transfer learning: What, how and when to transfer. What to transfer concerns the part of knowledge that is transferable across domains or tasks. On the one hand, the knowledge may be domain or task specific. On the other hand, it may be common across different domains and tasks. In the latter situation transfer learning may be beneficial. How to transfer concerns the learning algorithm for transfer learning, such as fine tuning, layer-wise freezing of an ANN or domain adaptation. Lastly, when to transfer also includes the question of when not to transfer to avoid a negative impact on the performance in the target domain which is known as negative transfer

2.2.1. Computer Vision

Transfer learning is well established in the field of computer vision. For example, for image classification pre-trained convolutional neural network (CNN) models like VGG-16, ResNet50, and Inceptionv3 are popular. These models were trained on large datasets, like the VGG-16 on 1.3 million images with 1,000 classes (Simonyan and Zisserman 2014).

Yosinski, Clune, Bengio, and Lipson (2014) stated that on the first convolutional layers CNNs learn abstract features such as edges and corners that are relevant for many datasets and tasks. When only a much smaller target dataset is available, learning such abstract features with a good generalization might be difficult. In this case, the first layers of the pre-trained CNN are useful as feature extractor and improve generalization. Building on these features, some of the higher convolutional layers and the classifier part of the CNN generate specialized features.

In addition, features from CNNs are suitable for many computer vision tasks (Razavian, Azizpour, Sullivan, and Carlsson 2014). The major advantage of transfer learning in computer vision is the elimination of the lengthy training process. Rawat and Wang (2017) provided a comprehensive overview of the application of CNNs to visual tasks.

2.2.2. Battery Ageing

Compared to the application of transfer learning in computer vision, pre-trained models related to battery ageing or SOH forecasting do not exist. This is potentially due to the fact that there are only few publicly available battery ageing datasets. In the field of battery ageing standardized tasks like RUL prediction and SOH estimation exist, just like image classification and image segmentation in computer vision. However, input features are very different regarding complexity and aggregation which complicates providing pre-trained models.

For transfer learning applied to SOH forecasting no direct related work exists. However, there exist two works in the context of transfer learning and battery ageing: Azkue, Lucu, Martinez-Laserna, and Aizpuru (2021) trained a baseline MLP using only calendar ageing data of 30 NMC cells. Then they apply transfer learning with a reduced amount of another LFP dataset. As benchmark they trained a model solely on the LFP dataset. They achieved a better performance with two LFP cells and transfer learning than without transfer learning and five LFP cells. Moreover, their results indicate improved generalization with transfer learning. However, they only compared to one benchmark model and use only one target dataset. In addition, the transfer is only examined for calendar ageing, not cyclic ageing.

Shen, Sadoughi, Li, Wang, and Hu (2020) trained a CNN for capacity estimation of lithium-ion batteries on a source dataset of ten years of cycling ageing data. All five convolutional layers and the three fully-connected layers of the pre-trained model were transferred. The best performance was achieved, when fine-tuning all layers providing the most capabilities for model adaption to the target domain. Compared to Azkue et al. (2021), Shen et al. (2020) vary the amount of target data available for fine-tuning. They find that training a benchmark model from scratch needs three times more training samples compared to transfer learning. Thus, transfer learning saves time and costs for data collection in their application.

In summary, these two papers indicate that transfer learning is a promising approach for battery ageing models. However none of these papers, considered the temporal sequence when data becomes available for training. This consideration is important to answer the question “when to transfer.” Further, each paper only used one target dataset. Thus, they cannot be used to compare the influence of different cell chemistry and battery operation on the success of transfer learning.

3. METHOD

In Section 3.1, a short overview of a method for SOH forecasting is given whose suitability for the task was shown (von Bülow et al. 2021). Building on this method for SOH forecasting, we introduce a training process for transfer learning in Section 3.2 to overcome the lack of training data for new battery types. This training process can also be applied to other machine learning methods in the field of battery ageing.

3.1. State of Health Forecasting

As mentioned in Section 2.1, battery ageing is perceived as a state change from a current $SOH(t_1)$ to a future $SOH(t_2)$ due

to ageing causes. The ageing causes are encoded in the battery operational data which consists of multidimensional time series signals of c-rate, temperature, and SOC. As depicted in Figure 2 first, this battery operational data is used to extract stressor table data of battery stressor types which are known to induce battery ageing. Second, the flattened stressor table data is input of a machine learning (ML) model, that outputs the state change ΔSOH from a current $SOH(t_1)$ to a future $SOH(t_2)$. The SOH values are assumed to be known for the training data. The two parts of the proposed SOH forecasting method are explained in detailed by von Bülow et al. (2021).

Figure 2: Model structure - stressor extraction (1) and ML model (2) (von Bülow et al. 2021)

3.2. Transfer Learning

First, a SOH forecasting MLP model in D_S, the source model, is (pre-)trained according to Section 3.1. Second, the source model is parametrically transferred to another domain D_T that has only little data available at transfer time.

We assume D_S to be a battery type with plenty of training samples available. Further, D_T represents a new battery type that has little available training samples at the point in time t of the transfer ($|D_T(t)| \ll |D_S|$). We assume that in time the amount of training samples of the new batteries increases ($|D_T(t_1)| < |D_T(t_2)|$ for $t_1 < t_2$). We preprocess the battery operational time series data according to von Bülow et al. (2021). For all datasets the same signal interval width and the same boundaries of the stressor tables in D_S and D_T are used so that $\mathcal{X}_S = \mathcal{X}_T$. However, for different operation and different battery types $P(X_S) \neq P(X_T)$ hold true because the distribution within the stressor tables and also $SOH(t_1)$ is different. As $P(X_S) \neq P(X_T)$, also the domains are different ($D_S \neq D_T$).

Regarding the learning tasks, we assume $\Upsilon_S = \Upsilon_T$ because we only output ΔSOH. However, we state that $f_S(\cdot) \neq f_T(\cdot)$ because the range of ΔSOH is different for the battery types in $\mathcal{T}_S$ and $\mathcal{T}_T$ given $X_S = X_T$. E.g. the target domain may be unbalanced with more new batteries. As $f_S(\cdot) \neq f_T(\cdot)$, also the tasks are different ($\mathcal{T}_S \neq \mathcal{T}_T$). As source and target task are different, we face the setting of inductive transfer learning. Thus, we need to use labeled data from the target domain to induce $f_T(\cdot)$ (Pan and Yang 2010).

3.2.1. What to Transfer: Common Characteristics

In the context of this work, "what to transfer" concerns the common characteristics of the relationship of current $SOH(t_1)$, battery operation (c-rate, temperature, and SOC) and the future $SOH(t_2)$.

3.2.2. How to Transfer: Layer Freezing

We accomplish knowledge transfer by a parametric transfer of the weights and biases from a source model to the target model. As common procedure in computer vision shown in Section 2.2.1, weights and biases of the model's layers can be frozen, i.e. they do not change while further training. Thus, the knowledge leant from D_S is preserved in the frozen layers. In the unfrozen layers, the weights pre-trained in D_S serve as starting point compared to randomly initialized weights. After freezing, we continue training the target model with the target training data. The parametric transfer has the transfer hyperparameter learning rate α, epochs, and number of frozen layers n_{frozen}. The number of frozen layers is counted from the input layer towards the output layer as the example in Figure 3 shows. For answering "how to transfer", we examine how to optimize n_{frozen} for achieving the best result on the target test data.

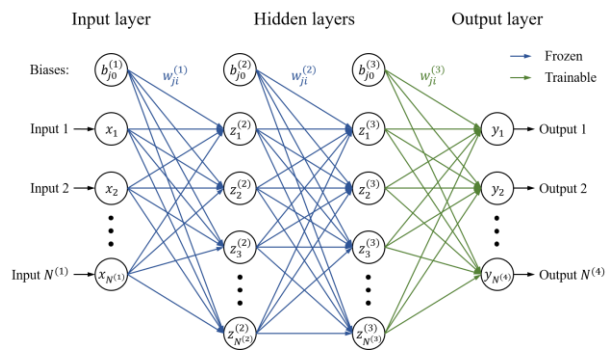


Figure 3: MLP with exemplary freezing of two layers (first and second). The output layer is trainable.

3.2.3. When to Transfer: Data Availability

"When to transfer" for battery ageing models refers to the temporal availability of target data. Thus, we artificially split the target data into training and validation data which is available at transfer time and test data which has not been recorded at transfer time. The test data will be input to the model when it runs in production, i.e. in prediction mode applied e.g. by automotive manufacturer or BEV fleet managers. Thus, model evaluation on the test data is a suitable measure for the success of the transfer. The amount of training target data is measured by the number of samples available for the transfer. The distribution is defined by the quantity and age of batteries, i.e. the maximum cycle number.

The target data is split according to three scenarios inspired by an automotive manufacturer that will introduce a BEV with a new battery type. The new battery type is the target domain D_T from a machine learning point of view. For each data split, first the order of the battery cells is shuffled. Then samples are added to the training and validation dataset according to the data split until the specified number of samples is reached. If multiple window lengths are present in the dataset, we sort ascendingly. The remaining samples

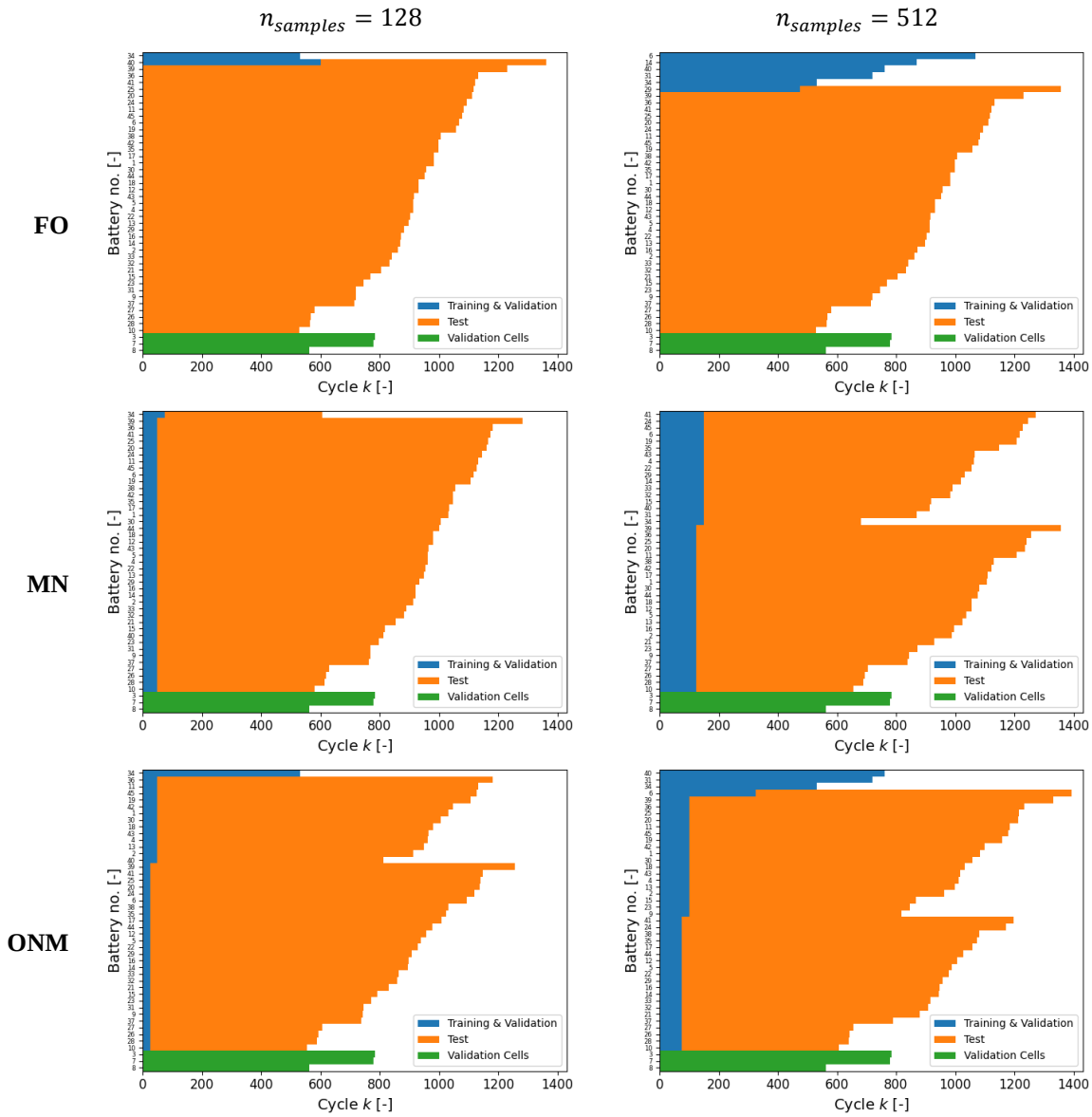


Figure 4: Exemplary data splits "few old" (FO), "many new" (MN), and "old new mixed" (ONM) with $n_{\text{samples}} = \{128; 512\}$ using the Closed-loop dataset and random seed 0.

compose the test dataset. The three data splits are shown exemplary in Figure 4:

First, the data split "few old" (FO) assumes that before the market launch of the new BEV type endurance tests are run: A few vehicles are intensively driven so that components like the batteries age quickly (Choi, Jung, Ham, and Bae 2011; Gassner 1984). Further, limited data from laboratory experiments is possibly available. Thus, target data of few, but old batteries is available for transfer learning.

Second, the data split "many new" (MN) assumes that after the market launch of a new BEV type after some time already many vehicles are driven by customers. However, these batteries have only aged very little or non-measurably. Thus, target data of many, but new batteries is available for transfer learning.

Third, the data split "old new mixed" (ONM) assumes the availability of data from the first two data splits. Thus, target data of many, but new batteries and few, but old batteries is available for transfer learning.

3.2.4. Benchmarks

For comparing the performance of transfer learning with non-transfer learning approaches whilst having the same availability, we define three benchmarks with a different composition of training and validation data as shown in Table 2. All benchmarks are tested on the same test data like the

Table 2: Overview of benchmarks for transfer learning

Benchmark	Training and validation data	Test data
Source Only (SO)	D_S	
Source Target Mixed (STM)	D_S & D_T	D_T
Target Only (TO)	D_T	

MLPs with transfer learning according to the data split type and the number of samples available. Only the training and validation data differ. Source Only (SO) is the source model without further modifications tested on the target test data. Source Target Mixed (STM) uses data from source and target domain for training and validation. The Target Only (TO) model is exclusively trained and validated on target domain data. For comparability, all benchmarks are MLPs with the same network architecture as the source model, but the weights are randomly initialized. Only their learning rate and batch size are optimized.

4. DATA

We selected public datasets that first provide data of sufficient battery cells, second have overall the same as well as different battery types, and third have a variety of battery operation to examine transfer learning in different circumstances. We differentiate battery types among others regarding materials and nominal capacity. Regarding battery operation charging, discharging, and storage of the batteries are relevant. This variety enables us to examine the effect of different battery types and operation. An overview of the used datasets regarding battery operation and materials is given in Table 3.

In the following, for each dataset we first present a brief description. Then, if necessary, we describe applied data cleansing methods, if physical inconsistencies in the data like time jumps or measurement errors were found. Lastly, cells that were not used and the reasons for this decision are specified.

The following two procedures were applied to more than one dataset:

- For the NASA Random and Oxford dataset, the start of a new cycle is not given. Thus, we define cycle starts when the current in a new step switches from zero or below zero (storage or discharging) to higher than zero (charging) so that every cycle starts with charging. But we require the previous cycle to be at least a full equivalent cycle (FEC) before a new cycles starts.
- For ISEA, NASA Random, and Oxford, capacity measurements are conducted in regular intervals. The SOH values are interpolated over time in between these capacity measurements.

4.1. Stanford Datasets

Severson et al. (2019) and Attia et al. (2020) present public datasets using the same battery cells with varying fast charging protocols. We refer to these two datasets by the paper's name first using the dataset as "Data-Driven" and "Closed-Loop" respectively. These commercial LFP/graphite cells, manufactured by A123 systems (APR18650M1A), were cycled in a forced convection temperature chamber set to 30°C under varied fast charging conditions but identical discharging conditions of 4C. The cells have a nominal capacity of 1.1 Ah. The sampling rate is 1s.

The SOH values are defined by the discharged electric charge of the corresponding cycle. The SOC signal is calculated offline relatively to the discharged electric charge of the previous cycle.[2] SOH values corresponding to single cycles are identified as outliers when they deviate more than three local standard deviations from the local mean within a 30-element window of the neighboring SOH values. These values are interpolated linearly using the neighboring SOH values.

4.1.1. Data-Driven

The Data-Driven (DD) dataset (Severson et al. 2019) used in this work consists of 42 lithium-ion batteries belonging to the batch of 2017-05-12 cycled up to 80% SOH. All cells are charged with a two-step fast-charging protocol. This protocol has the format "C1(Q1)-C2", in which C1 and C2 are the C-rates of the first and second CC steps (CC1 and CC2 respectively) and Q1 is the SOC at which the current switches. C1 and C2 range from 3 to 8C, while Q1 ranges from 15 to 80 % SOC.[3] The second current step ends at 80% SOC, after which the cells charge with another CC step at 1C (CC3) followed by a CV phase.

The SOH signal is smoothed using a moving mean with a span of 15 data points because the SOH signal is noisy. The cells 1 to 5 of batch 2017-05-12 continued as cells 8 to 17 in batch 2017-06-30. Thus, we concatenated them in batch 2017-05-12. The thermocouples of cells 15 and 16 in batch 2017-05-12 were switched. This means that the temperature signals of these cells are not synchronized with the other cells' signals. It was not possible for us to synchronize them. Thus, we do not use these cells. Neither do we use cells no. 1 and 19 as some of their cycles are up to 11.8 times longer as the previous cycles because the switching from charging to discharging happens late. This would skew the min-max-normalization. We use cells no. 3, 7, and 8 as validation cells which means that they are not part of the training, validation or test set.

[2] The discharged electric charge is specified by the end value of the signal "Qd".

[3] C1 and C2 ∈ {3, 3.6, 4, 4.4, 4.8, 5.4, 6, 7, 8}C. Q1 ∈ {15, 25, 30, 35, 40, 50, 60, 70, 80} % SOC.

4.1.2. Closed-Loop

The Closed-Loop (CL) dataset (Attia et al. 2020) used in this work consists of 45 lithium-ion batteries belonging to the batch of 2019-01-24 cycled up to 22.8 % SOH. All cells are charged with a three-step fast-charging protocol. This protocol has the format "CC1-CC2-CC3", in which CC1, CC2, and CC3 are the C-rates of the first, second, and third CC steps that end at 20%, 40%, and 60% respectively. CC1, CC2 and CC3 each range from 3.6 to 8C. A fourth parameter, CC4, is dependent on CC1, CC2, CC3, and the charging time. CC4 ranges from 2.68 to 4.755C.[4]

The SOH signal is smoothed using a moving mean with a span of 25 data points because the SOH signal is noisy. We use cells no. 3, 7, and 8 as validation cells which means that they are not part of the training, validation or test set.

4.2. ISEA

The RWTH ISEA (ISEA) (Sauer 2021) battery dataset consists of 48 commercial lithium-ion battery cycled with the same profile under equal conditions to explore intrinsic cell manufacturing variability and small temperature differences in battery packs during operation. These cylindrical

Table 3: Overview of the used datasets and their relevant characteristics

Dataset		Stanford Data-driven (DD)	Stanford Closed-Loop (CL)	ISEA	NASA randomized (random)	Oxford
Source		(Severson et al. 2019)	(Attia et al. 2020)	(Sauer 2021)	(Bole, Kulkarni, and Daigle 2012)	(Raj, Wang, Monroe, and Howey 2020)
Battery ageing	Number of cells	42	45	48	26	27
	Life time range [cycles]	532-2,235	530-1,231	1,410-1,872	199-1,188	300-1,681
	Min. SOH [%]	80-88	22-76	50	50-86	2-86
Battery type	Nominal capacity [Ah]	1.1	1.1	≅1.85	2.1	3
	Nominal voltage [V]	3.3	3.3	3.7	?	3.6
	Voltage limits [V]	2.0-3.6	2.0-3.6	3.0-4.1	3.2-4.2	2.5-4.2
	Packaging style	18650 (cylindrical)	18650 (cylindrical)	18650 (cylindrical)	18650 (cylindrical)	18650 (cylindrical)
	Cathode material	LFP	LFP	NMC	LCO	NCA
	Anode material	Graphite	Graphite	Carbon	Graphite	Graphite
Battery operation	Max. charging c-rate	3-8C	3.6C-8C	2.2C	2.2C	0.5C
	Max. discharging c-rate	-4C	-4C	-2.2C	-2.4C	-0.5C
	ΔSOC	100%	100%	60% (SOC 20-80%)	Up to 100%	100 %
	Ambient temperature [°C]	30	30	25	20 & 40	24
	Storage time	0h	0h	0h	0h	5 or 10 d

[4] CC1, CC2, and CC3 ∈ {3.6, 4.4, 4.8, 5.2, 5.6, 6, 7, 8}C. CC4 ∈ {2.68, 3, 3.652, 3.834, 3.94, 4.16, 4.252, 4.755}C.

NMC/graphite cells, manufactured by Panasonic/ Sanyo (UR18650E), were cycled at constant ambient temperature of 25°C. One cycle consists of 30 min discharging to 3.5 V and 30 min charging to 3.9 V, both currents limited to a maximum of 4 A. The charge turnover is about 1 Ah corresponding to cycles between approximately 20% and 80% SOC. Because of the voltage limits the charge turnover varies with the SOH of the cell, but the DOD in relation to the aged capacity is being kept nearly constant. The cells were graded into group C from the manufacturer and are drawn from the same production lot. Due to this factory selection the cells only have a mean capacity of approximately 1.85 Ah which we chose as nominal capacity. Capacity measurements are on average conducted every 165 cycles. The sampling rate is on average 0.1s.

The SOC signal calculation is orientated at the given boundaries of approximately 20% and 80% SOC for each cycle. Jumps of the time signal where corrected for cell 3 and 8. If the temperature signal is 0°C for a complete cycle, the ambient temperature of 25°C was set because this is the lowest physically possible temperature. This ensures that these states are considered in the stressor tables at least in the ambient temperature bin. If the temperature dropped to 0°C only for some time stamps, it was interpolated using the neighboring values because sudden temperature breaks are seen as impossible as the temperature only changes slowly. Cell 47 has a noisy temperature signal which was smoothed using an Fourier transform (FT) and moving mean. For some cells the last cycle length is 1,820 min instead of 60 min. For cell 29 and 18 the last 3 and 86 cycles respectively are very short. Because of these irregularities, we only use data corresponding to SOH higher than 50%.

4.3. NASA Randomized

NASA Ames Prognostics Center of Excellence Randomized Battery Usage Data Set (NASA Random) (Bole, Kulkarni, and Daigle 2012) consists of 26 battery cells of type LCO/graphite LG Chem. 18650 that were cycled in seven groups with different cycling protocols for each group (Details Appendix Table 7). The cycling protocols specify randomized sequences of current loads ranging from 0.5 A to 4 A. The sequences are randomized in order to better represent practical battery usage. The temperature was environmentally controlled. The sampling rate is 1s. After every fifty randomized discharging cycles, the capacity was determined (reference discharge cycles). These capacity measurements are executed twice.

Some capacity measurements were incomplete, resulting in a lower capacity value. Thus, if the capacity values deviate more than 0.1 Ah from each other, the higher values is chosen. The SOC signal calculation is orientated at the voltage limits after charging and discharging. We identified voltage jumps from 3.2 to 4.2 V combined with a time jump which are likely due to missing charging data. For best correction we added values with the sampling rate of 1s of the dataset: The current so that battery is fully loaded with CC. Temperature and voltage are linearly interpolated. Else the average of both is used. Batteries RW2 and RW18 were not used because the temperature signal has many values of -4,000°C which lies below absolute zero.

4.4. Oxford

Oxford Path Dependent Battery Degradation Dataset (Oxford) (Raj, Wang, Monroe, and Howey 2020) consists of 27 battery cells of type Panasonic NCA/graphite 18650 that were cycled in four groups with different cycling protocols for each group. All cells were cycled with CC between 0% and 100% SOC. After a period of cycling, calendar ageing was performed at 90% SOC. The time ratio of cyclic to calendar ageing was 1:5 with different c-rates and storage SOC (Details Appendix Table 6). Compared to the other datasets, only the Oxford dataset contains calendar aging. The temperature was environmentally controlled at 24°C. Capacity measurements are conducted every 48 cycles. The sampling rate is 1s.

The SOC signal calculation is orientated at the voltage limits after charging and discharging. Again, we justify the replacement of temperatures below 0°C and Not a Number values (NaN) by the ambient temperature of 24°C, as we aim to minimize adulteration of the training data ($T_{\min} = 24°C$, $T_{\max} = 36.6°C$, $\emptyset T = 25.2°C$). Inconsistency like negative time deltas and time jumps of “TestTime” higher than the sampling rate with following constant time values were corrected by replacing the time signal with the sampling rate of 1s onwards. This sampling rate was determined from the consistent time signals. This adaption seems legitimate to us, especially for the time jumps, because the duration of the time jumps equals the product of sampling rate and the amount of following constant time values. The file no. 27 of battery number 9 from part 2, group 1 has some single current values around -10^{282} A during CC phases. These were replaced by the average of the two neighbor current values. Battery no. 1 from part 2, group 6 was not used because it is only exposed to calendar ageing which prohibits the identification of cycles. Battery no. 10 from part 1, group 2 was not used because it only contains a single cycle. The file no. 14 and 15 of battery no. 14 from part 1, group 3 are renamed and used for battery no. 15 of the same part and group because these are the only files of battery no. 14. At the same time exactly these file numbers are missing for battery no. 15.

5. Results and Discussion

5.1. Design of Experiments

We aim at answering when and how to transfer a SOH forecasting model from the source to the target domain. For developing the parametric transfer the essential model

hyperparameters are the learning rate α and the number of frozen layers n_{frozen}. In addition, we compare the transfer learning to the benchmarks proposed in Section 3.2.4. Further, we examine the number of samples and their distribution over different batteries and ageing states (data split type). Model users are also interested in how the number of samples and data split type are influenced by the similarity of source and target dataset.

The target datasets for training and validation have 128, 256, 512, 768, and 1024 samples. This is limited by the NASA Random dataset which has only 1183 samples. We use a learning rate in the common range of 0.001, 0.0001, and 0.0001. Training did not converge with higher learning rates. We use the RMSE as evaluation metric because it is common for regression problems and, compared to the MSE, has the same unit as the predicted output value.

The source dataset of all experiments is the DD dataset which was also used in previous works (von Bülow et al. 2021). The source dataset contains 3127 training samples and each 391 validation and test samples. The source model is obtained following the method presented in previous work using the DD dataset again (von Bülow et al. 2021). The DD dataset provides an ideal starting point for our experiments because of its similarity to CL. We use fine 2D stressor table, variant A because this showed the best result in previous work (Appendix Table 4 and Table 5). Further, we assume that the fine signal interval width enables the model to learn the target data better. Batteries in the Oxford and NASA random dataset have a lifetime below 400 cycles. This prohibits window lengths of 400 and 530 (W5 and W6 in previous work) as well as the corresponding grouped datasets W10 to W12. Further, a grouped window length of {25; 50; 500} cycles (named W9 in previous work) showed promising results regarding generalization in the source domain. Thus, we chose W9 and a window shift of $w_{\text{s}} = 25$. The hyperparameters, model complexity and model performance of the source model on the source data are shown in the Appendix in Figure 9.

For better comparability, the benchmark MLPs use the same amount of layers and neurons as the source model. For TO and STM, only the learning rate $\alpha \in \{10^{-N} | 3 \leq N \leq 5\}$ and the batch size $bs \in \{16; 32; 64; 128\}$ are hyperparameters of a grid search. For SO, the source model is simply run in prediction mode as it is already trained only with the source data.

Compared to previous work (von Bülow et al. 2021), the min-max normalization is adapted to be suitable for the transfer learning. In previous work, min-max normalization was executed feature-wise using the source training data, i.e. each feature separately was normalized based on its min and max value in the source training data. The minima in the source and target data are mostly zero because the dwell time for every stressor type is zero. However, the max values of the target training data are different as stated in Section 3.2 ($P(X_S) \neq P(X_T)$). Furthermore, determining the maxima of the features of the test target data is not possible at transfer time, because the test target data becomes only available in the future. Thus in this work, min-max normalization is executed in two groups: First, for the $SOH(t_1)$ and second for all other features, i.e. the stressor values from the stressor tables. We use the maximum of all stressor values of the source training data for normalization. Further, we keep the same scaling from the source data for all target data to ensure that the weight of the stressor values is the same for all features. In addition, this eases the model's transfer because the model does not need to adapt to a different normalization of the target data.

Experiment 1: How to transfer? Freezing & Learning rate

In general, we expect n_{frozen} and α to jointly influence the convergence speed and also the final target model performance. n_{frozen} is a measure how much information from the source domain is kept at transfer time. Thus, we have a special interest in it as it also is a measure of similarity between source and target domain. We assume that a frozen model will perform well, if the domains are similar. Contrarily, if less information on the future ageing is given in the target domain training data, freezing is also beneficial to prevent negative transfers.

Experiment 2: When to transfer? Data splits & no. of samples

After examining the selection of the best model regarding n_{frozen} and α in Experiment 1, the question "when to transfer" concerns the suitability of the target data specified by the three data splits FO, MN, and ONM as well as the number of samples.

Experiment 2a:Closed-Loop

As visible in Table 3, the batteries in DD and CL have the same battery characteristics. Only the battery operation deviates regarding the charging protocols. Thus, DD and CL are similar domains, providing an ideal situation for transfer learning.

Experiment 2b: ISEA, NASA random, and Oxford

The transfer to the ISEA dataset provides a change in cathode material from LFP to NMC and a smaller SOC range. Batteries in automotive applications like BEVs experience variable discharging and not simple CC discharging like batteries in laboratory operation (von Bülow and Meisen 2022). Compared to the other datasets, in the NASA random dataset not CC discharging, but stepwise randomly changing discharge rates are applied. Not only cyclic ageing, but also calendar ageing is important in automotive applications for battery ageing which is considered by the Oxford dataset.

Because of these differences that influence the stressor data ($\subset P(X_T)$) and the predictive function $f_T(\cdot)$, we expect these

three datasets to be more challenging for successful transfer learning than the CL dataset. Furthermore, the NASA random and Oxford datasets are a step towards applying the SOH forecasting model and its transfer to BEV operational ageing data.

Each non-test target dataset was randomly split for training and validation by the ratio of 90:10. The validation MSE was set as metric for early stopping with patience of 5 epochs and minimum delta of 0 to avoid overfitting on the training target data. We apply a maximum of 90 epochs. The constant hyperparameters from source model training are not changed: Optimizer Adam, MSE as loss function, and a linear activation function of the output layer. We also opted for keeping the kernel regularizer for the transfer to prevent overfitting. Version 2.8.0 of TensorFlow was used as back-end including version 2.8.0 of Keras.

5.2. Evaluation of Experiments

When selecting a model regarding the hyperparameters n_{frozen} and α for a given target dataset, the validation and the test target RMSE are possible selection criteria. Both have a different distribution of the output values because of the temporal data split specified in Section 3.2.3. which leads to different ageing rates for samples from the beginning of life (BOL) compared to the end of life (EOL) (see Appendix Figure 9). Due to these assumptions of data availability for the model selection at transfer time only the validation target set is available, even though we are overall interested in the test target RMSE which indicates the model's performance when run in production.

Result experiment 1: How to transfer? Freezing & Learning rate

In experiment 1 we do not only evaluate which n_{frozen} and α are optimal in the majority of scenarios, but we also evaluate the reduction of model performance on the test dataset due to the selection criteria only being the validation RMSE. Exemplarily, Figure 5 shows the model's RMSE of all three α with different n_{frozen} for ISEA, MN with 128 samples. Selecting by the test RMSE (rectangles) $\alpha = 0.001$ and $n_{\text{frozen}} = 7$ would be optimal, but by the validation RMSE (crosses) $\alpha = 0.0001$ and $n_{\text{frozen}} = 0$ is best.

As depicted in Figure 6 for all four target datasets and selected by the validation RMSE, freezing no layer is the best choice in 95 % of the data splits ($n_{\text{frozen}} = 0$). However, when selecting by the test RMSE this would only be the case in 45 % of the data splits. 18 % of all data splits have a better test RMSE with freezing the first layer instead of none. This indicates that in the majority of the data splits no general features have been learnt from the source domain. Compared to the state of the art procedure in the field of computer vision this is potentially due to the limited amount of source training data. $n_{\text{frozen}} = 10$ selected by the test RMSE is only optimal to prevent a negative transfer in the case of MN and CL as further discussed in Experiment 2.

Further, for α the selection based on validation and target RMSE diverges even more as shown in Figure 7. For 72 %

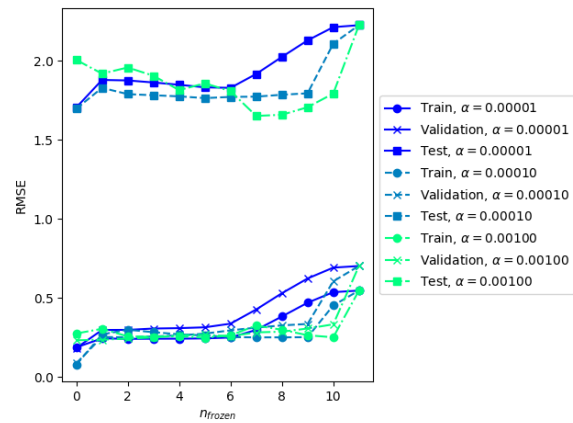


Figure 5: Exemplary model selection (ISEA, MN, 128 samples)

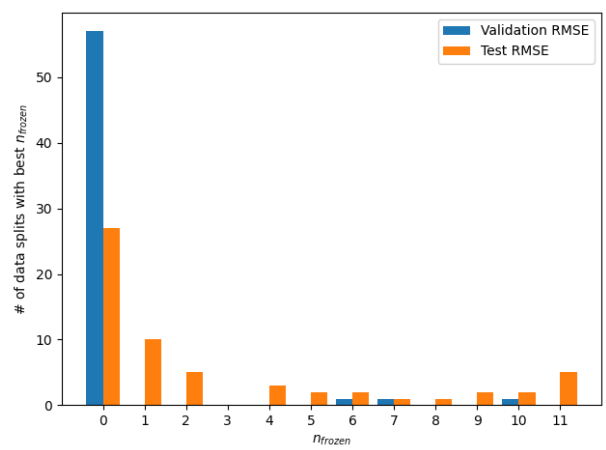


Figure 6: Best n_{frozen} selected by validation and test RMSE of all four target datasets

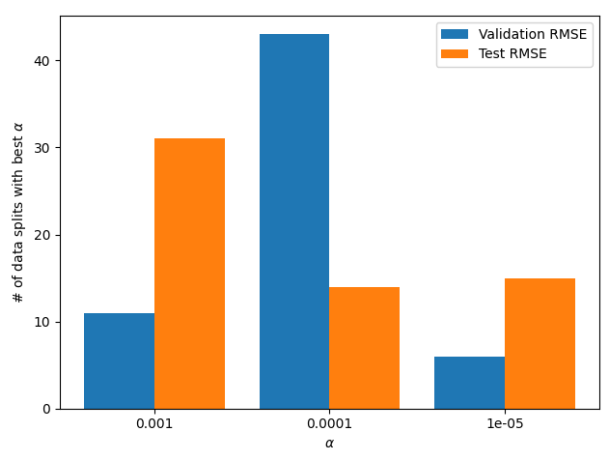


Figure 7: Best learning rate α selected by validation and test RMSE of all four target datasets

of the data splits 0.0001, which is also the source model's α, is the best α when selected by the validation RMSE, but for 60% of the data splits 0.001 is the best when using the test RMSE. Overall, these discrepancies lead to an average deviation of the test RMSE when selecting $n_{\text{frozen}}$and α by the validation RMSE of 23% with a standard deviation of 28 %. Unfortunately, we have to accept this error because we cannot access the test RMSE itself for model selection. In the following, for each data split we select the best model based on the validation RMSE.

Experiment 2: When to transfer? Data splits & sample no.

Figure 8 depicts the performance of the best model selected by validation RMSE on training, validation, and test data as well as the SO, TO, and STM benchmarks of all four target datasets. We first analyze CL as target dataset and then the remaining three target datasets.

Experiment 2a: Closed-Loop

Figure 8a) shows for the CL target dataset FO that the training and validation RMSE are on the same level as the source RMSE (see Appendix Table 8 for comparison). The test RMSE is higher overall and just acceptable for an application of the model e.g. by fleet managers who are interested in forecasting degradation from 100% down to approximately 80%. The test RMSE is decreasing once more data becomes available for training. This is coherent with our expectations. Independently of the amount of data only STM can compete with transfer learning. TO only reaches a similar model performance with 1,024 training samples.

For MN, we observe a negative transfer that leads to a RMSE of around 5 which is too high for a practical model application. This is caused by a lack of degradation information about the knee point. Even with 1,024 samples not sufficient information about ageing at EOL and the knee point is provided because 1,024 samples only correspond to the first 250 cycles and a minimum SOH of 93%, but the earliest knee point of CL is at around 500 cycles (Appendix Figure 9).

For MN, also the benchmarks SO and STM perform better than transfer learning and TO is as bad as transfer learning. This underlines the unsuitability of the data of MN for transfer learning.

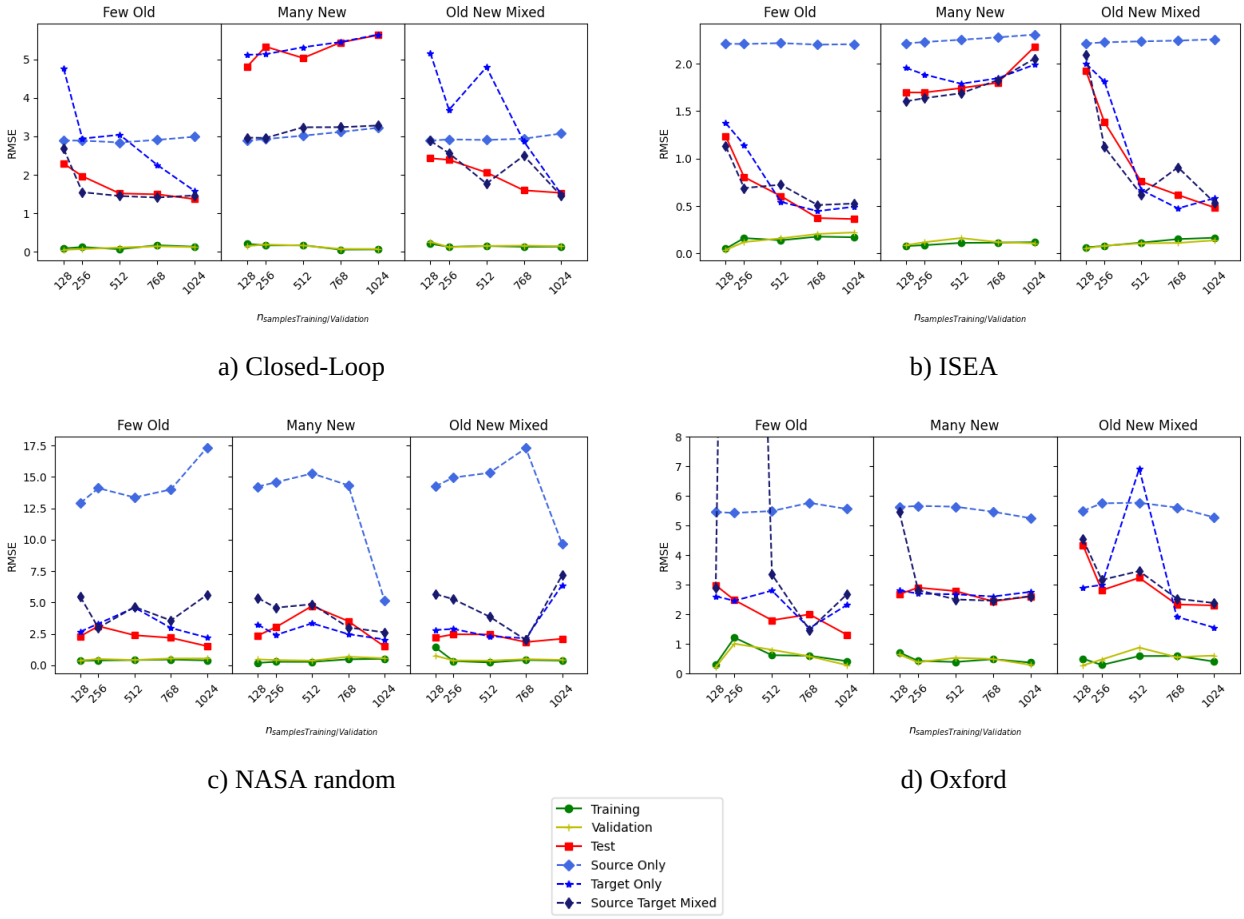


Figure 8: Best models selected by validation RMSE of all four target datasets and their benchmarks

For the ONM, the model performance is similar to FO. Thus adding MN to FO does not provide additional relevant information to the model. Again TO performs worse than the transferred model.

Experiment 2b: ISEA, NASA random, and Oxford

The results using the target datasets ISEA, NASA random, and Oxford confirm that FO is the preferred scenario of data availability. Only for NASA random and Oxford the test RMSE of FO and MN is similar, but still FO is better. We assume that this is caused by a constant ageing rate of NASA random and Oxford (see Appendix Figure 9). Contrarily, DD, CL, and ISEA show superlinear degradation and, thus, have a knee point. This knee point makes forecasting with the MN data split more difficult than when having linear degradation because data of the accelerated ageing after the knee point is missing. Consequently, the test RMSE of ONM of CL and ISEA is better than of MN because ONM contains information about the knee point.

Overall, the SO benchmark is worse than the other benchmarks and transfer learning except in 16% of the cases. This indicates that any data from the target dataset improves the model's performance. The comparison of the TO and STM benchmarks with transfer learning leads to no unambiguous results as each of the three is performing slightly better in some cases. Unfortunately, we could not identify any clear pattern under which circumstances which of the three is preferable. Probably, a more extensive source dataset is required for a clear added value of transfer learning. One could for example, use four of the presented datasets as source dataset and only one as target dataset.

6. Conclusion

This paper showed under which conditions transfer learning enables SOH forecasting based on a comprehensive study using four known public target battery datasets: Having data of few aged batteries shall be preferred over data of many young batteries, especially in the case of superlinear degradation with knee points. In the case of data of few aged batteries, having data of more aged batteries from the target domain will improve model performance. We also discussed the problem of an available metric (Validation vs. test RMSE) which prohibits selecting the best possible model.

In contrast to state of the art transfer learning in computer vision, freezing no layers was best for the transfer with most datasets and data splits. This is due to the limited amount of source data and coherent to the results of Shen et al. (2020).

The results are biased by the MLP architecture chosen by a hyperparameter optimization based on the source dataset. Furthermore, none of the dataset originates from BEV operation, but only laboratory data was used. Only the NASA random and Oxford dataset come close to BEV operational data with variability in discharge c-rate and calendar ageing respectively. Thus, transfer learning from laboratory to BEV operational data is planned as future work. In that case, possibly another advantage of transfer learning might become relevant if models are trained on thousands of samples from BEV fleets. Like in computer vision, then transfer learning might save usage of computational resources when pre-training a model requires weeks of training, but transfer learning significantly less.

In this work, we only considered a single battery type each in the source and target domain. Now that conditions for transfer learning enabling SOH forecasting are known, future work could examine SOH forecasting for several battery types whose data become available sequentially. This paradigm is called continual learning. Continual learning is similar to transfer learning, but aims not only at transferring to a single new task, but to several sequentially occurring tasks. Thus, catastrophic forgetting after the first new task becomes relevant for continual learning (Parisi, Kemker, Part, Kanan, and Wermter 2019). Continual learning in the context of SOH forecasting is imaginable in two scenarios: First, with progressing time more and more data of the same battery type or fleet will be available. Each time, integrating this new data into an existing model is a task of continual learning. Second, a single SOH forecasting model may be desirable for several battery types of different generations that are only developed in temporal sequence. Once, data of a new battery type is accessible an existing model shall be made suitable for the old battery types and the new one.

Connected to continual learning, curricular learning is worth examining to find a training strategy that presents training samples not randomly, but organized in a meaningful order. This enables the model to gradually learn more complex concepts (Bengio, Louradour, Collobert, and Weston 2009).

Acknowledgement

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. The results, opinions and conclusions expressed in this publication are not necessarily those of Volkswagen Aktiengesellschaft.

References

Attia, P. M., Grover, A., Jin, N., Severson, K. A., Markov, T. M., Liao, Y.-H., Chen, M. H., Cheong, B., Perkins, N., Yang, Z., Herring, P. K., Aykol, M., Harris, S. J., Braatz, R. D., Ermon, S., & Chueh, W. C. (2020). Closed-loop optimization of fast-charging protocols for batteries with machine learning. *Nature*, vol. 578 (7795), pp. 397–402. doi: 10.1038/s41586-020-1994-5

Attia, P. M., Bills, A. A., Brosa Planella, F., Dechent, P., dos Reis, G., Dubarry, M., Gasper, P., Gilchrist, R., Greenbank, S., Howey, D., Liu, O., Khoo, E., Preger, Y., Soni, A., Sripad, S., Stefanopoulou, A., & Sulzer,

V. (2022). Review—"Knees" in Lithium-Ion Battery Aging Trajectories. *Journal of The Electrochemical Society*, vol. . doi: 10.1149/1945-7111/ac6d13

Azkue, M., Lucu, M., Martinez-Laserna, E., & Aizpuru, I. (2021). Calendar Ageing Model for Li-Ion Batteries Using Transfer Learning Methods. *World Electric Vehicle Journal*, vol. 12 (3), p. 145. doi: 10.3390/wevj12030145

Barré, A., Deguilhem, B., Grolleau, S., Gérard, M., Suard, F., & Riu, D. (2013). A review on lithium-ion battery ageing mechanisms and estimations for automotive applications. *Journal of Power Sources*, vol. 241 (9), pp. 680–689. doi: 10.1016/j.jpowsour.2013.05.040

Bengio, Y., Louradour, J., Collobert, R., & Weston, J. (2009). Curriculum learning. *Proceedings of the 26th Annual International Conference on Machine Learning (ICML)*, vol., pp. 41–48. doi: 10.1145/1553374.1553380

Bertoldi, N., Cettolo, M., Federico, M., & Buck, C. (2012). Evaluating the Learning Curve of Domain Adaptive Statistical Machine Translation Systems. *Proceedings of the 7th Workshop on Statistical Machine Translation Montréal, Canada, June 7-8, 2012*, vol., pp. 433–441.

Bole, B., Kulkarni, C. S., & Daigle, M. (2012). *Randomized Battery Usage Data Set*. Available at: https:// ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/ (last accessed February 4, 2021).

Chen, L., Lü, Z., Lin, W., Li, J., & Pan, H. (2018). A new state-of-health estimation method for lithium-ion batteries through the intrinsic relationship between ohmic internal resistance and capacity. *Meausrement*, vol. 116, pp. 586–595. doi: 10.1016/j.measurement.2017.11.016

Choi, Y., Jung, D., Ham, K., & Bae, S. (2011). A study on the accelerated vibration endurance tests for battery fixing bracket in electrically driven vehicles. *Procedia Engineering*, vol. 10, pp. 851–856. doi: 10.1016/j.proeng.2011.04.140

Fermín-Cueto, P., McTurk, E., Allerhand, M., Medina-Lopez, E., Anjos, M. F., Sylvester, J., & dos Reis, G. (2020). Identification and machine learning prediction of knee-point and knee-onset in capacity degradation curves of lithium-ion cells. *Energy and AI*, vol. 1 (5), p. 100006. doi: 10.1016/j.egyai.2020.100006

Gassner, E. (1984). Vereinfachter Betriebsfestigkeits-Nachweis für zufallsartig im hohen Lebensdauerbereich beanspruchte Fahrzeugbauteile / Time-reduced performance fatigue test for automotive components randomly loaded in the high endurance range. *Materials Testing*, vol. 26 (8), pp. 274–276. doi: 10.1515/mt-1984-260807

Gewald, T., Candussio, A., Wildfeuer, L., Lehmkuhl, D., Hahn, A., & Lienkamp, M. (2020). Accelerated aging characterization of lithium-ion cells. Using sensitivity analysis to identify the stress factors relevant to cyclic aging. *Batteries*, vol. 6 (1), p. 6. doi: 10.3390/batteries6010006

Keil, P. (2017). *Aging of Lithium-Ion Batteries in Electric Vehicles*. PhD Thesis, Technische Universität München, München, Germany, https://mediatum.ub.tum.de/doc/1355829/file.pdf.

Leuthner, S. (2018). Lithium-ion battery overview, InR. Korthauer (Ed.), *Lithium-Ion Batteries: Basics and Applications* (pp. 13–19). Berlin, Heidelberg: Springer Berlin Heidelberg.

Lipu, M. S. H., Hannan, M. A., Hussain, A., Hoque, M. M., Ker, P. J., Saad, M. H. M., & Ayob, A. (2018). A review of state of health and remaining useful life estimation methods for lithium-ion battery in electric vehicles. Challenges and recommendations. *Journal of Cleaner Production*, vol. 205, pp. 115–133. doi: 10.1016/j.jclepro.2018.09.065

Marongiu, A., Roscher, M., & Sauer, D. U. (2015). Influence of the vehicle-to-grid strategy on the aging behavior of lithium battery electric vehicles. *Applied Energy*, vol. 137, pp. 899–912. doi: 10.1016/j.apenergy.2014.06.063

Matadi, B. P., Geniès, S., Delaille, A., Waldmann, T., Kasper, M., Wohlfahrt-Mehrens, M., Aguesse, F., Bekaert, E., Jiménez-Gordon, I., Daniel, L., Fleury, X., Bardet, M., Martin, J.-F., & Bultel, Y. (2017). Effects of biphenyl polymerization on lithium deposition in commercial graphite/NMC lithium-ion pouch-cells during calendar aging at high temperature. *Journal of The Electrochemical Society*, vol. 164 (6), A1089-A1097. doi: 10.1149/2.0631706jes

Nuhic, A., Terzimehic, T., Soczka-Guth, T., Buchholz, M., & Dietmayer, K. (2013). Health diagnosis and remaining useful life prognostics of lithium-ion batteries using data-driven methods. *Journal of Power Sources*, vol. 239 (3), pp. 680–688. doi: 10.1016/j.jpowsour.2012.11.146

Palmer, J. A. (2003). *Relative Convexity*. Available at: https://sccn.ucsd.edu/~jason/relcon_new.pdf (last accessed December 16, 2021).

Pan, S. J., & Yang, Q. (2010). A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering*, vol. 22 (10), pp. 1345–1359. doi: 10.1109/TKDE.2009.191

Parisi, G. I., Kemker, R., Part, J. L., Kanan, C., & Wermter, S. (2019). Continual lifelong learning with neural

networks: A review. *Neural Networks*, vol. 113, pp. 54–71. doi: 10.1016/j.neunet.2019.01.012

Raj, T., Wang, A. A., Monroe, C. W., & Howey, D. A. (2020). Investigation of Path‐Dependent Degradation in Lithium‐Ion Batteries**. *Batteries & Supercaps*, vol. 3 (12), pp. 1377–1385. doi: 10.1002/batt.202000160

Rawat, W., & Wang, Z. (2017). Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review. *Neural Computation*, vol. 29 (9), pp. 2352–2449. doi: 10.1162/neco_a_00990

Razavian, A. S., Azizpour, H., Sullivan, J., & Carlsson, S. (2014). CNN Features Off-the-Shelf: An Astounding Baseline for Recognition. *2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (pp. 512–519), 2014-06-23/2014-06-28, Columbus, OH, USA. doi: 10.1109/CVPRW.2014.131

Richardson, R. R., Osborne, M. A., & Howey, D. A. (2019). Battery health prediction under generalized conditions using a Gaussian process transition model. *Journal of Energy Storage*, vol. 23, pp. 320–328. doi: 10.1016/j.est.2019.03.022

Ruder, S., Peters, M. E., Swayamdipta, S., & Wolf, T. (2019). Transfer Learning in Natural Language Processing. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Tutorials* (pp. 15–18), Minneapolis, Minnesota. doi: 10.18653/v1/N19-5004

Sauer, D. U. (2021). *Time-series cyclic aging data on 48 commercial NMC/graphite Sanyo/Panasonic UR18650E cylindrical cells*. Available at: https://doi.org/10.18154/RWTH-2021-04545 (last accessed July 12, 2021).

Severson, K. A., Attia, P. M., Jin, N., Perkins, N., Jiang, B., Yang, Z., Chen, M. H., Aykol, M., Herring, P. K., Fraggedakis, D., Bazant, M. Z., Harris, S. J., Chueh, W. C., & Braatz, R. D. (2019). Data-driven prediction of battery cycle life before capacity degradation. *Nature Energy*, vol. 4 (5), pp. 383–391. doi: 10.1038/s41560-019-0356-8

Shao, L., Zhu, F., & Li, X. (2015). Transfer learning for visual categorization: a survey. *IEEE Transactions on Neural Networks and Learning Systems*, vol. 26 (5), pp. 1019–1034. doi: 10.1109/TNNLS.2014.2330900

Shen, S., Sadoughi, M., Li, M., Wang, Z., & Hu, C. (2020). Deep convolutional neural networks with ensemble learning and transfer learning for capacity estimation of lithium-ion batteries. *Applied Energy*, vol. 260 (9), p. 114296. doi: 10.1016/j.apenergy.2019.114296

Simonyan, K., & Zisserman, A. (2014). *Very Deep Convolutional Networks for Large-Scale Image Recognition*. Available at: http://arxiv.org/pdf/1409.1556v6.

Sulzer, V., Mohtat, P., Aitio, A., Lee, S., Yeh, Y. T., Steinbacher, F., Khan, M. U., Lee, J. W., Siegel, J. B., Stefanopoulou, A. G., & Howey, D. A. (2021). The challenge and opportunity of battery lifetime prediction from field data. *Joule*, vol. 5 (8), pp. 1934–1955. doi: 10.1016/j.joule.2021.06.005

Torrey, L., & Shavlik, J. (2010). Transfer Learning, In E. S. Olivas, J. D. M. Guerrero, M. Martinez-Sober, J. R. Magdalena-Benedito, & A. J. Serrano López (Eds.), *Handbook of Research on Machine Learning Applications and Trends* (242-264). Hershey, PA, USA: IGI Global.

Viering, T., & Loog, M. (2021). *The Shape of Learning Curves: a Review*. Available at: http://arxiv.org/pdf/2103.10948v1.

von Bülow, F., Heinrich, F., & Meisen, T. (2021). Fleet Management Approach for Manufacturers displayed at the Use Case of Battery Electric Vehicles. *2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (pp. 3218–3225), 2021-10-17/2021-10-20, Melbourne, Australia. doi: 10.1109/SMC52423.2021.9658680

von Bülow, F., Mentz, J., & Meisen, T. (2021). State of health forecasting of Lithium-ion batteries applicable in real-world operational conditions. *Journal of Energy Storage*, vol. 44, p. 103439. doi: 10.1016/j.est.2021.103439

von Bülow, F., & Meisen, T. (2022). A Review on Methods for State of Health Forecasting of Lithium-Ion Batteries applicable in Real-World Operational Conditions. *Unpulished*, vol. .

von Srbik, M.-T. (2015). *Advanced lithium-ion battery modelling for automotive applications*. PhD Thesis, Imperial College London, London.

Vuorilehto, K. (2018). Materials and function, InR. Korthauer (Ed.), *Lithium-Ion Batteries: Basics and Applications* (pp. 21–28). Berlin, Heidelberg: Springer Berlin Heidelberg.

Waag, W., Fleischer, C., & Sauer, D. U. (2014). Critical review of the methods for monitoring of lithium-ion batteries in electric and hybrid vehicles. *Journal of Power Sources*, vol. 258, pp. 321–339. doi: 10.1016/j.jpowsour.2014.02.064

Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). How transferable are features in deep neural networks? *27th International Conference on Neural Information*

Processing Systems (NIPS) (pp. 3320–3328), 2014-12-08/2014-12-13, Montreal, Canada.

APPENDIX

Table 5: Signal interval width for current, temperature, and SOC

	Signal interval width		
	I	**T**	**SOC**
Fine (F)	0.5C	0.5 °C	5 %
Medium (M)	1C	1 °C	10 % at 0 and 100%, else 20 %
Coarse (C)	3C	3 °C	20 %

Table 4: Combined signals for 2D stressor tables, variant A

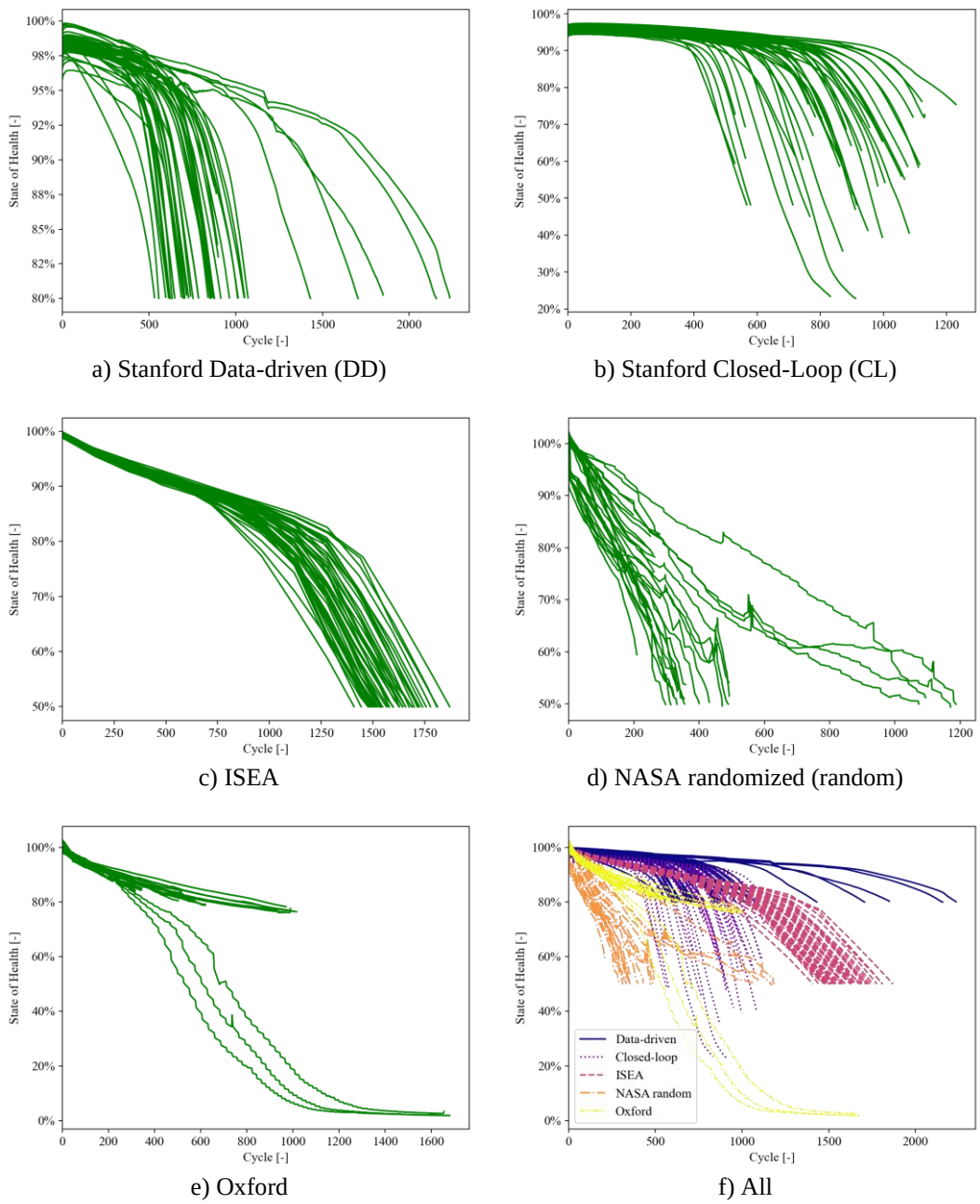
	Charging	**Discharging**	**Hold**
variant A	T & SOC I & SOC I & T	T & SOC I & SOC I & T	T & SOC

Table 6: Overview of battery operation Oxford dataset

Dataset part	**Group no.**	**Cell no.**	**Cycling**			**Calendar aging**	
			Duration [d]	**C-rate**	**Cycling type**	**Duration [d]**	SOC_{Storage}
1 & 2	1	9, 15, 20	1	C/2	CC	5	90%
	2	3, 4, 8		C/4			
	3	10, 11, 14	2	C/2		10	
	4	12, 18, 19		C/4			
2	5	5, 6, 16	From 2.5V-4.2V	C/2	CC	-	-
	6	1	-	-	-		90%
3	7	4, 19, 27	1	C/2	CCCV	5	90%
	8	5, 18, 20	2			10	90%
	9	1, 8, 9	1			5	4.2V
	10	2, 22, 25				10	4.2V

Table 7: Overview of battery operation NASA random dataset. *CC(CV): CC then CV if enough time

Group no.	Cell no.	Charging			Discharging			$T_{Ambient}$ [°C]
		Duration	C-rate	Cycling type	C-rate	Length random sequence	High probability	
1	9-12	5 min or U=4.2V	Rand. up to 4.5A	CC(CV)*	Rand. up to 4.5A	5 min or U=3.2V	-	20
2	3-6	To 4.2V	2A	CCCV	Rand. 0.5-4A	5 min	-	20
3	1,2,7,8	Rand. 0.5 - 3 h	C/2	CC(CV)*			-	20
4	25-28	To 4.2V	2A	CCCV	Rand. 0.5-2A	1 min	High *I*	40
5	17-20							20
6	21-24						Low *I*	40
7	13-16							20



a) Stanford Data-driven (DD)

b) Stanford Closed-Loop (CL)

c) ISEA

d) NASA randomized (random)

e) Oxford

f) All

Figure 9: SOH over cycle number *k* of all used datasets. Discontinuities appear because of plotting over cycles, but the SOH is interpolated over time for ISEA, NASA Random, and Oxford.

Table 8: Hyperparameters of the source model

		{25, 50, 100} cycles (W9)
Hyperparameters		
Activation Function		ReLU
Batch Size		64
Learning Rate α		0.0001
Regularization $\{\lambda_1, \lambda_2\}$		{0, 0.001}
Dropout rate		0
MLP layout		[195, 45, 395, 95, 245, 295, 245, 145, 245, 245]
Model Complexity		
No. of Hidden Layers		10
No. of Model Parameter		1,693,126
Metrics		
RMSE	**Train**	0.0861
	Validation	0.1032
	Test	0.1083
$\mathbf{R^2}$	**Train**	0.9967
	Validation	0.9961
	Test	0.9935