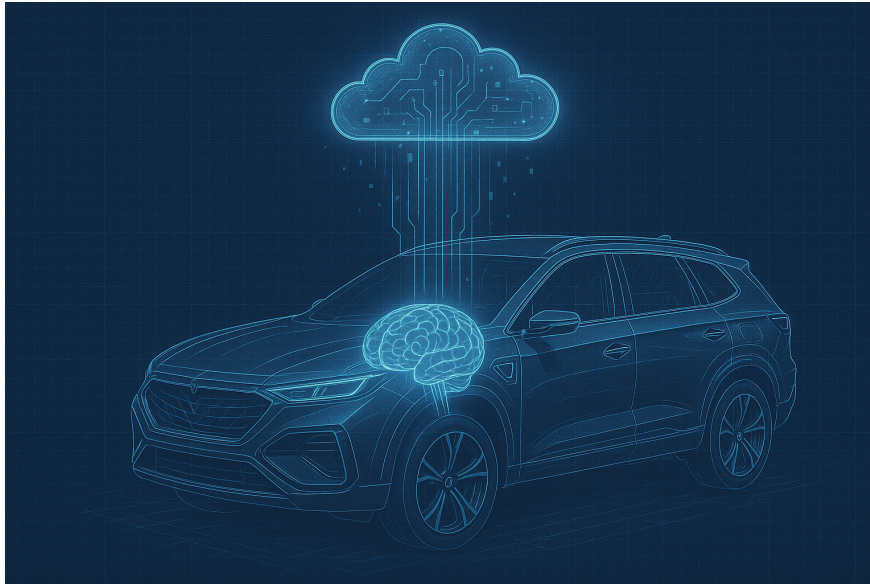




CHALMERS
UNIVERSITY OF TECHNOLOGY



Transfer Learning for Battery Health Forecasting: From Lab to Real-World Data

Personalized Degradation Models for Electric Vehicle Batteries

Master's thesis in Complex Adaptive Systems

Oskar Andersson, Ludvig Fornstedt

DEPARTMENT OF ELECTRICAL ENGINEERING

CHALMERS UNIVERSITY OF TECHNOLOGY

Gothenburg, Sweden 2025

www.chalmers.se

MASTER'S THESIS 2025

Transfer Learning for Battery Health Forecasting: From Lab to Real-World Data

Personalized Degradation Models for Electric Vehicle Batteries

OSKAR ANDERSSON, LUDVIG FORNSTEDT



CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Electrical Engineering
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2025

Personalized Battery Management Systems
Through Decentralized Data-Driven Learning
Personalized Degradation Models for Electric Vehicle Batteries
OSKAR ANDERSSON, LUDVIG FORNSTEDT

© OSKAR ANDERSSON, LUDVIG FORNSTEDT, 2025.

Supervisor: Christian Fleischer, Cognivity AI AB
Supervisor: Xiaolei Bian, Systems and Control, Electrical Engineering
Examiner: Changfu Zou, Systems and Control, Electrical Engineering

Master's Thesis 2025
Department of Electrical Engineering Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Typeset in L^AT_EX
Printed by Chalmers Reproservice
Gothenburg, Sweden 2025

Personalized Battery Management Systems Through
Decentralized Data-Driven Learning
Personalized Degradation Models for Electric Vehicle Batteries
OSKAR ANDERSSON, LUDVIG FORNSTEDT
Department of Electrical Engineering
Chalmers University of Technology

Abstract

Accurately forecasting lithium-ion battery health in electric vehicles remains challenging due to the scarcity and variability of real-world data and the disconnect between controlled laboratory tests and in-service operation. To address this, a transfer-learning framework is proposed, that leverages diverse lab datasets and small amounts of vehicle-specific data to produce personalized State of Health (SOH) and Remaining Useful Life (RUL) forecasts. The proposed method employs a dual LSTM architecture, where one branch ingests historical SOH trajectories, while a parallel branch processes simple statistical descriptors (mean and standard deviation of voltage, current, and temperature) per cycle. The outputs of the two LSTMs are concatenated and passed through a lightweight MLP to yield cycle-wise forecasts. Models were trained on three open-source lab datasets (MIT, XJTU, HKUST) encompassing varied chemistries and cycling protocols, then evaluated both on a held-out lab domain and on real-world EV data from nine vehicles spanning 18–30 months of operation. Results demonstrate that the dual LSTM consistently outperforms simpler baselines, with fine-tuning on early-life data yielding substantial accuracy gains. Our framework thus effectively provides a step towards bridging the lab-to-road gap, enabling scalable, adaptive battery management.

Keywords: RUL, SOH, LSTM, Battery health, Battery Degradation, Transfer Learning, EV

Acknowledgments

We would first like to extend our thanks to our supervisor at Cognivity AI, Christian Fleischer. Christian has provided both great guidance and support during the process of this master thesis. We would also like to extend our thanks to Xialoei Bian and Changfu Zou, our supervisors at E2. They have provided input, guidance and given us insights from contemporary research.

Oskar Andersson, Ludvig Fornstedt, Gothenburg, June 2025

List of Acronyms

Below is the list of acronyms that have been used throughout this thesis listed in alphabetical order:

BMS	Battery management system
BOL	Beginning of life
EOL	End of life
EV	Electric Vehicle
LSTM	Long Short-Term Memory
MAPE	Mean average percentage error
MOL	Middle of life
RUL	Remaining Useful Life
SOC	State of Charge
SOH	State of Health

Contents

List of Acronyms	ix
Nomenclature	xi
List of Figures	xiii
List of Tables	xv
1 Introduction	1
1.1 Previous work	2
1.1.1 SOH estimation of Li-ion batteries	2
1.1.2 RUL prediction of Li-ion batteries	3
1.2 Transfer learning of RUL forecasts between battery domains	4
1.3 Problem description and research aim	4
2 Theory	7
2.1 Battery degradation	7
2.2 Long short-term memory models (LSTM)	8
3 Data and Methodology	11
3.1 Data	11
3.1.1 Data Description	11
3.1.2 Lab Data Processing	13
3.2 Feature construction and engineering	13
3.3 Model approaches and architectures	15
3.4 Final Dual stream LSTM architecture	16
3.5 Evaluation	17
3.5.1 Evaluation on lab data	17
3.5.2 Evaluation on EV data	18
4 Results and Discussion	19
4.1 Feature analysis	19
4.2 Model performance on lab data	20
4.3 Transfer learning results on EV data	22
5 Conclusion	23
Bibliography	25

List of Figures

2.1	Internal structure of a standard LSTM cell. The cell state runs horizontally through the memory cell, while the input, forget, and output gates regulate the flow of information.	9
3.1	Unfiltered capacity trajectories of MIT and XJTU data sets.	12
3.2	Comparison of the different LSTM architectures which are examined. In this thesis the main input consists of State of health data and the extra inputs consists of the features discussed in 3.2.	15
4.1	Spearman correlations between features for both lab-data sets.	20
4.2	RMSE scores for the models when evaluated on different data sets. The plots to the left highlight the model performance when trained and validated on batteries with identical operating conditions. The plots to the right highlight the performance when transferred to a new domain.	21
4.3	Comparison of RMSE values and improvements for different starting cycles.	22
4.4	Model evaluation of the different retraining modes on the EV-data for the final dual-LSTM model.	22

List of Tables

3.1	Summary of Lab Datasets	12
3.2	Summary of Selected Features	14

1

Introduction

The advent of widespread usage of electrical vehicles (EVs) has brought high expectations on performance. One of the leading issues with EVs is the life expectancy of the batteries powering them. The problem here lies within the complex aging behavior of lithium-ion batteries and the unpredictability of their lifespan; as well as their changing behavior depending on environmental factors such as temperature. Battery degradation is significantly influenced by temperature extremes, cycling frequency, depth of discharge, and charge/discharge rates, all of which exacerbate the complexity and variability of battery aging [1, 2].

To ensure both safe and reliable usage of the batteries, accurate State of Health (SOH) and Remaining Useful Life (RUL) estimations are needed. SOH is commonly defined as the ratio of the battery's present capacity to its rated nominal capacity and serves as an indicator of degradation severity [1, 2]. RUL indicates how many more cycles the battery is useful for, usually defined as the remaining cycles until battery capacity degrades below 80% of its original rated capacity, marking the practical end-of-life threshold [3]. This is important from both a user standpoint, to be able to tell how much mileage is left in the car, and from a safety standpoint, to know the actual health of the battery.

Current research is attacking this problem from many directions where the main focus is either in understanding the complex aging mechanisms of batteries, or in developing data-driven methods for accurate SOH estimations. Model-based methods, including electrochemical and equivalent circuit models, provide physical insight but can be computationally intensive and require detailed parameterization, limiting their generalizability to real-world applications [4]. Data-driven methods, by contrast, have demonstrated increased flexibility and accuracy in capturing complex, nonlinear aging behaviors using real-world operational data [5, 6]. Among data-driven methods, neural network-based approaches such as Long Short-Term Memory (LSTM) models have emerged as particularly powerful due to their ability to capture long-range temporal dependencies inherent to battery degradation patterns [4, 5, 6].

A smaller part of this research is dedicated to the development of predictive RUL models. This is due in large part to the difficulty of accurately predicting long-term degradation patterns under variable conditions. As a result, robust and generalizable

RUL prediction remains an open challenge in the field. There are also few studies which focus on battery data from real-world EVs rather than less relevant battery data from laboratory settings. Due in large part to the scarcity of labeled data, significantly restricting the availability of comprehensive training data.

To address these limitations, this thesis explores data-driven techniques for SOH and RUL forecasting, with particular emphasis on leveraging transfer learning and neural sequence models to bridge the gap between laboratory datasets and real-world electric vehicle operation. Transfer learning has recently been introduced to battery health estimation as a promising method to generalize laboratory-based models to real-world battery applications, significantly reducing the volume of real-world data required for accurate model training and adaptation [7, 8, 4]. Neural sequence models, particularly LSTM networks, have demonstrated exceptional capabilities for temporal modeling of battery degradation, making them a promising tool for accurate battery health forecasting under realistic EV operating conditions [3, 8, 4].

1.1 Previous work

The following section provides an overview of the current landscape in battery health modeling. The discussion is divided into three parts. First, common methods for SOH estimation are outlined, including both model-based and data-driven approaches. This is followed by a review of predictive frameworks for SOH forecasting and RUL estimation, which remain less developed but increasingly important. Finally, recent work on transfer learning is examined, with a particular focus on efforts to adapt models trained on laboratory data to real-world electric vehicle operation.

1.1.1 SOH estimation of Li-ion batteries

Accurate estimation of a lithium-ion battery's SOH is essential to ensure performance reliability, safety, and efficient lifecycle management in electric vehicles (EVs). SOH is commonly defined as the ratio of the battery's present capacity to its rated nominal capacity and is used as an indicator of degradation severity [1, 2]. Estimating SOH, however, remains a non-trivial challenge due to the complex, nonlinear, and highly variable nature of battery aging under real-world operational conditions [9].

Approaches to SOH estimation can broadly be categorized into three groups: model-based, direct measurement-based, and data-driven methods. Model-based techniques, including equivalent circuit models (ECMs) and physics-informed electrochemical models, attempt to describe internal battery behavior through theoretical constructs. While these models can be informative and interpretable, they often require detailed knowledge of the cell's internal parameters and may struggle to generalize across different chemistries, usage conditions, or battery pack configurations [10, 11, 4].

Direct measurement approaches rely on metrics and methods such as internal resis-

tance and Coulomb counting. While potentially accurate, these methods are often intrusive or impractical to implement continuously within on-board systems, especially under dynamic load profiles or fleet-scale variability [4].

In contrast, data-driven SOH estimation methods have gained significant attention due to their scalability, adaptability, and compatibility with real-time applications. These methods use historical and real-time sensor data, such as current, voltage, temperature, and state of charge (SOC), to learn relationships between observable battery signals and degradation indicators. A wide range of algorithms has been applied, from classical machine learning methods such as support vector regression and random forests to modern deep learning techniques. These approaches eliminate the need for intrusive measurements or in-depth electrochemical modeling, making them attractive for large-scale deployment [6, 5].

Among deep learning methods, recurrent architectures, particularly Long Short-Term Memory (LSTM) networks, have emerged as a leading choice for SOH estimation. LSTMs are well-suited for time-series modeling, as they are capable of capturing both short-term fluctuations and long-range dependencies in sequential data. This makes them especially effective in battery applications, where degradation patterns often evolve gradually across many charging and discharging cycles. LSTM-based models have shown high predictive accuracy, even under variable cycling conditions, and have demonstrated robustness when trained with limited datasets [4, 6, 5].

The ensemble model proposed by Che et al. exemplifies the effectiveness of LSTM networks when integrated with other deep learning architectures. Their framework combines DNN, CNN, and LSTM models to forecast SOH in series battery packs. The LSTM component is specifically used to model the temporal degradation patterns, and its inclusion significantly improves generalization under heterogeneous load conditions [4, 10]. Their results show that LSTM-based predictors maintain high accuracy even when trained on as little as 10–30% of the total data, highlighting the model’s strength in low-data regimes.

Similarly, Xu et al. demonstrated improved SOH estimation using a CNN-LSTM hybrid architecture, allowing the network to first extract local features through convolutional layers before passing temporal patterns to the LSTM module [5]. Such hybrid models have consistently outperformed traditional methods by leveraging both spatial and temporal representations of battery behavior.

1.1.2 RUL prediction of Li-ion batteries

While SOH estimation has seen substantial progress in both theoretical development and industrial deployment, the prediction of RUL of lithium-ion batteries remains comparatively less explored. RUL refers to the estimated duration, often expressed in charge-discharge cycles or operating hours, until a battery reaches its defined end-of-life (EOL) threshold, typically when capacity falls below 80% of the rated value. In contrast to SOH, which provides a snapshot of current battery health,

RUL forecasts the trajectory of degradation and is therefore central to long-term planning, maintenance scheduling, and second-life deployment strategies [3, 8, 12].

Like SOH estimation, RUL prediction techniques can be broadly categorized into model-based and data-driven approaches. Model-based methods, such as particle filters and state-space models, simulate internal degradation mechanisms through mathematical representations of chemical and physical processes [4, 8]. However, these methods often suffer from limited generalizability and sensitivity to parameter tuning.

In similarity to the research on SOH models, LSTM architectures have been proven efficient and reliable in RUL forecasting research. Recent studies propose LSTM-based models for RUL prediction that demonstrate strong performance even with relatively small training sets and uncertain test conditions [12]. The leading method for RUL forecasting is an iterative approach, as seen in [13, 14, 3], where a sliding window is applied to the degradation time series.

1.2 Transfer learning of RUL forecasts between battery domains

Transfer learning frameworks within RUL forecasting is sought after because of its potential to generalize between battery types and cycling behaviors. As with SOH and RUL forecasting, LSTM remain the dominating model type here [8, 15], but with varying methods for the frameworks. The difficulty of transfer learning is the large change in forecasting domain between battery types, cycling behavior and data gathering. Contemporary research is searching for transfer learning for both different cycling behavior, [16, 17] and for different battery types [3]. Retraining often takes place for a specific held out battery, where the model is then validated on similar batteries; or retrained for the first cycles of a battery from a different domain before forecasting.

1.3 Problem description and research aim

Battery health estimation in EVs presents a unique set of challenges, driven by the complex and variable nature of battery degradation. This variability arises from a combination of chemical, mechanical, and thermal processes that evolve over the course of a battery’s life, influenced by diverse operating conditions, usage patterns, and environmental factors [1, 2, 8]. Real-world data on battery degradation is often sparse, noisy, and incomplete, complicating efforts to develop accurate, generalizable models for State of Health (SOH) and Remaining Useful Life (RUL) prediction [7, 4]. The scarcity of available real-world-data further exacerbates this issue.

While laboratory data provides detailed, controlled measurements of battery performance, these data sets are rarely representative of the wide-ranging conditions encountered in actual EV fleets. This gap between controlled lab data and real-

world operational data presents a significant challenge for battery health modeling, as models trained solely on laboratory data often fail to generalize to practical applications without further adaptation [8, 6].

To address these challenges, this thesis aims to develop a flexible, data-driven framework for battery health estimation that can effectively leverage both laboratory and real-world data. The approach focuses on integrating personalized, vehicle-specific features with global, fleet-wide trends to improve the scalability and generalization of battery health models. Transfer learning is explored as a key method for bridging this gap, allowing laboratory-trained models to adapt to the diverse and often noisy conditions of real-world EV fleets without requiring extensive additional data [7, 8].

2

Theory

The following section introduces the necessary theory to effectively discuss the given problem and subsequent solutions. First battery degradation and its mechanisms will be discussed, to give some understanding to the underlying process of the given problem. Secondly, the neural network models which were utilized will be covered.

2.1 Battery degradation

Battery degradation in electric vehicles (EVs) is a critical concern as it directly impacts vehicle performance, range, and overall reliability. Lithium-ion batteries, which dominate EV applications due to their high energy density, are subject to two primary degradation pathways: calendar aging and cycling aging. Calendar aging refers to the loss of battery capacity and increased internal resistance occurring over time, even without battery cycling, predominantly due to chemical reactions within the battery cells. In contrast, cycling aging results from the continuous process of charging and discharging, where the battery electrodes and electrolyte undergo physical and chemical changes leading to decreased performance [1].

Several factors accelerate these degradation mechanisms, including temperature extremes, depth of discharge (DOD), charge and discharge rates, and mechanical stresses within battery cells. High temperatures particularly accelerate aging by intensifying chemical reactions, causing electrolyte decomposition, Solid Electrolyte Interphase (SEI) growth, and cathode material instability. Similarly, low temperatures hinder ionic conductivity and promote lithium plating, further contributing to performance degradation and potential safety risks. The SEI layer, initially protective, becomes thicker and less conductive over time, impeding lithium-ion transport and increasing internal resistance. Additionally, lithium plating, especially under conditions of high charging currents and low temperatures, not only reduces battery efficiency but also poses severe safety hazards due to dendrite formation, which can cause internal short circuits. Battery degradation significantly manifests in reduced State of Health (SOH), characterized by decreased capacity and power relative to the battery's original specifications [2].

2.2 Long short-term memory models (LSTM)

Sequential data appears across numerous domains, ranging from natural language processing and financial forecasting to control systems and time-series prediction. In such contexts, learning dependencies over time is essential. Traditional Recurrent Neural Networks (RNNs) were designed to model such temporal dependencies by maintaining a hidden state that evolves with each time step. However, RNNs have proven inadequate for learning long-range dependencies due to the vanishing and exploding gradient problems during training, which severely limits their performance on extended sequences. The Long Short-Term Memory (LSTM) architecture was introduced by Hochreiter and Schmidhuber in 1997 to directly address these limitations [18].

LSTM models enhance traditional RNNs through the introduction of memory cells and a gating mechanism that enables the network to learn when to store, update, or discard information. The core idea is to maintain a cell state that runs through the sequence with minimal linear interaction, allowing gradients to flow unimpeded through long time steps. Each LSTM unit contains three gates: the input gate, the forget gate, and the output gate. These gates are composed of sigmoid-activated neural layers that control the flow of information into and out of the cell state. The input gate determines which values will be updated, the forget gate decides what information is discarded, and the output gate regulates the output passed to the next time step. Figure 2.1 shows the general structure of the LSMT cell and its information flows.

An LSTM cell maintains a cell state C_t that acts as a long-term memory, which is updated at each time step through the interactions of the three main gates. The governing equations for the LSTM are written down below.

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \quad (2.1)$$

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \quad (2.2)$$

$$\tilde{C}_t = \tanh(W_C x_t + U_C h_{t-1} + b_C) \quad (2.3)$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \quad (2.4)$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \quad (2.5)$$

$$h_t = o_t \odot \tanh(C_t) \quad (2.6)$$

The forget gate f_t determines which portions of the previous cell state C_{t-1} should be retained, using a sigmoid activation function (eq. 2.1). Next, the input gate i_t controls which new information should be added to the cell state (eq.2.2). Then the candidate cell state \tilde{C}_t is generated as (eq. 2.3). The cell state is then updated by combining the previous cell state, modulated by the forget gate, and the candidate cell state, modulated by the input gate (eq. 2.4). Finally, the output gate o_t determines the contents of the hidden state h_t , which is passed to the next LSTM cell or output layer (eq. 2.5 - 2.6). Where W and U are the weight matrices, b are

the bias terms, and \odot denotes element-wise multiplication. The sigmoid activation function σ ensures that the gate values are constrained between 0 and 1, while the hyperbolic tangent function scales the candidate cell state to the range of -1 to 1, promoting stable training dynamics [19].

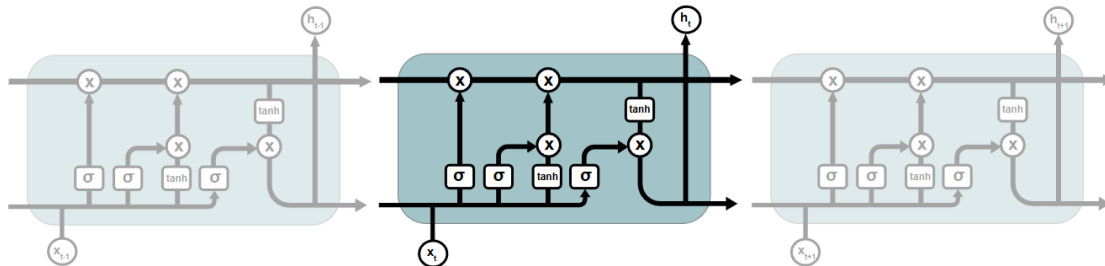


Figure 2.1: Internal structure of a standard LSTM cell. The cell state runs horizontally through the memory cell, while the input, forget, and output gates regulate the flow of information.

As stated, the inclusion of these gating mechanisms allows LSTMs to learn relationships over significantly longer time intervals than traditional RNNs. This makes them well-suited for tasks where past information is crucial for current prediction. For example, in natural language processing, LSTM networks have been used for tasks such as machine translation, language modeling, and speech recognition, where the context established by earlier words or sounds strongly influences the current output [20]. Similarly, in finance and economics, LSTMs are leveraged to predict stock prices, demand forecasting, and anomaly detection due to their robustness in handling noisy and non-stationary sequences [21].

In contrast to feedforward neural networks, where input-output pairs are assumed to be independent and identically distributed (i.i.d.), LSTM networks inherently assume sequential dependence. This makes them a natural choice in modeling dynamic systems where temporal evolution is central to the problem [22].

3

Data and Methodology

3.1 Data

The data for this thesis is split into two main groups, lab data and EV data. The lab data was gathered from several open-source data repositories, including the MIT[9], XTJU[23] and HKUST[24] datasets. These datasets consist of battery cycling data, including time, voltage, current, temperature and capacity measurements. A summary of the lab-data sources is provided in Table 3.1. The approach taken to data gathering was to find data sets which contain detailed SOH, cell temperature, voltage and current data. For the sake of transfer learning it was also deemed necessary for the data to encompass different charging and discharging modes. The EV data is provided by an undisclosed EV-manufacturer, spanning nine cars with varying operating lifetimes and usage conditions.

3.1.1 Data Description

The MIT Battery Dataset, introduced by Severson et al. [9], contains data from 124 commercial lithium iron phosphate (LFP) 18650 cells cycled under controlled laboratory conditions. The cells were tested in a temperature-controlled environment at 30°C, with a standardized 4C discharge rate and varied fast-charging protocols designed to produce a wide range of degradation trajectories. The dataset covers a total of 72 distinct fast-charging protocols, each designed to stress the cells differently, resulting in a wide distribution of cycle lives ranging from approximately 150 to 2,300 cycles [9].

The XJTU Battery Dataset, curated by Wang et al. [23], consists of 55 commercial nickel-cobalt-manganese (NCM) 18650 cells manufactured by LISHEN. These batteries were subjected to a variety of charging and discharging strategies designed to simulate diverse real-world usage conditions. This dataset is notable for its inclusion of six different cycling protocols, providing a more heterogeneous training set for machine learning models. Batteries were typically cycled until their capacity fell below 80% of the initial value, producing a diverse set of degradation trajectories [23].

Lastly, the HKUST data set, created by Tang et al. [24], consists of 215 lithium ion battery cells which were cycled under similar conditions. Due to its homogeneity,

this data set will be used as a test set within the context of this thesis.

Table 3.1: Summary of Lab Datasets

Dataset	Battery Type	Cycling Modes	No. of Batteries
MIT [9]	LFP/graphite	3	124
XJTU [23]	NCM	6	55
HKUST [24]	NCM	1	215

These datasets provide a diverse foundation for the training and evaluation of data-driven battery health estimation models, with the MIT data emphasizing high-stress, fast-charging conditions, and the XJTU data offering a broader range of cycling behaviors typical of real-world applications. The combination of these data sources is intended to support the development of robust, transferable machine learning models capable of accurately predicting battery degradation across a wide range of operating conditions. The SOH trajectories for the data sets are plotted in figure 3.1 below.

The real-world data was provided by an undisclosed car manufacturer. The data spans nine cars with operating data spanning from 18 to 30 months with largely middle of life (MOL) data. Since the EV batteries do not reach end of life (EOL), the usage of this data will only be introduced as a final validation that the models can adapt to real-world data.

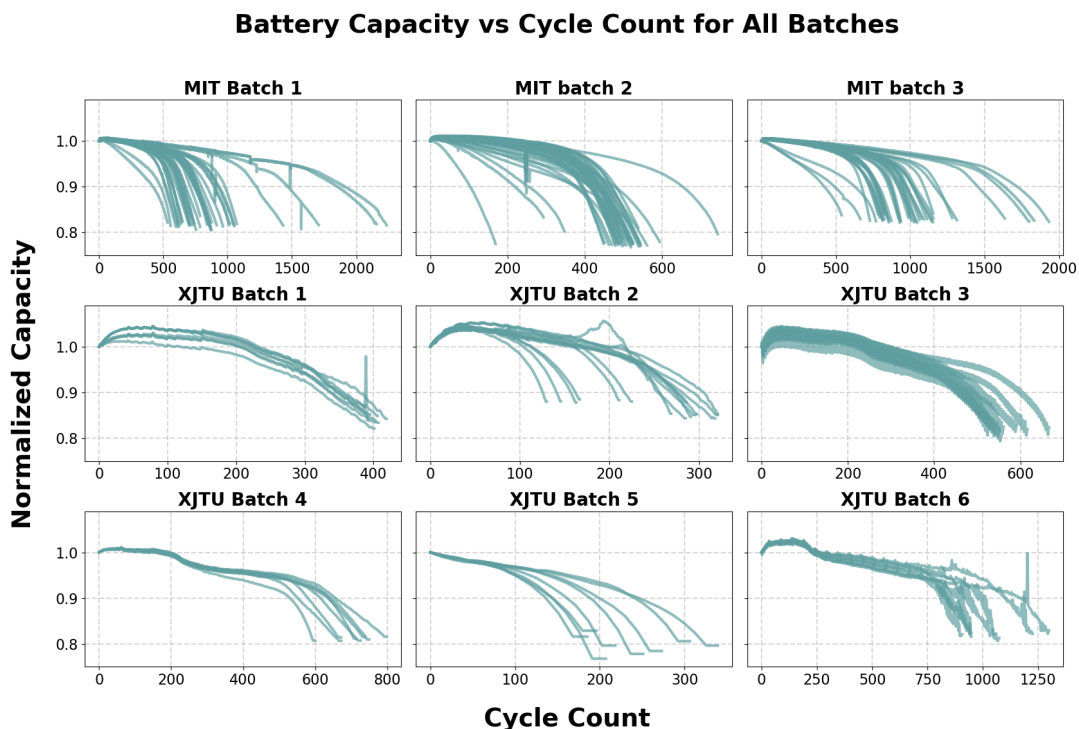


Figure 3.1: Unfiltered capacity trajectories of MIT and XJTU data sets.

3.1.2 Lab Data Processing

The lab data is initially stored in various formats, depending on the original repository. To standardize the data for training, all datasets undergo the same preprocessing pipeline, ensuring consistency in format and feature extraction across sources. This pipeline is composed of several steps beginning with data cleaning where noisy channels and incomplete cells are removed. For example, the MIT dataset includes several cells that do not reach 80% capacity or contain significant measurement noise, which are filtered out to improve model robustness. Then key features such as voltage, current, temperature, and capacity are extracted from the raw time-series data for each cycle. This is done by parsing the nested dictionary structure, extracting features like ‘V’, ‘I’, ‘T’, and capacity for each individual cycle. The processed data is then serialized into a standardized ‘pkl’ format, which preserves the hierarchical structure while reducing file size and loading time during training. This conversion allows efficient access to cycle-level data without the overhead of parsing raw text or matrix files.

3.2 Feature construction and engineering

This thesis has adopted a straightforward and interpretable approach to feature engineering. Given the complex nature of battery aging and the wide variability of operational conditions, feature extraction plays a critical role in effectively training predictive models. A core objective was to leverage the strengths of the LSTM architecture, which is capable of learning temporal dependencies directly from sequential data without the need for heavily processed input signals [18, 20, 8]. However, preliminary experiments revealed that feeding completely unprocessed time-series data resulted in computational inefficiency and model overfitting, largely due to the high dimensionality and inherent noise in raw measurements.

Many previous studies have employed physically motivated features such as incremental capacity (IC) curves, differential voltage (dQ/dV) analysis, or charge time at fixed voltage thresholds [9, 11]. While these features can be highly informative under controlled laboratory conditions, they are difficult to reproduce reliably in real-world settings due to incomplete data, variable charging protocols, and lack of full-cycle measurements. For the purpose of the thesis, it is necessary that the model gets the same feature quality from both domains.

Consequently, the decision was made to simplify the input signals while retaining sufficient information to capture key battery degradation patterns. Basic statistical descriptors, namely the mean and standard deviation, were computed from the raw voltage, current, and temperature measurements collected at each battery cycle. These simple statistical summaries intended to preserve the essential dynamics of battery operation while maintaining compatibility with the transfer learning and generalization objectives of this study. A summary of these features is presented in Table 3.2.

To investigate relationships among the constructed input features and their rele-

Table 3.2: Summary of Selected Features

Feature	Description
Voltage (mean)	Average voltage during a cycle
Voltage (std)	Standard deviation of voltage during a cycle
Current (mean)	Average current during a cycle
Current (std)	Standard deviation of current during a cycle
Max Cell Temp. (mean)	Average temperature during a cycle
Max Cell Temp. (std)	Standard deviation of temperature during a cycle

vance to battery degradation, a correlation analysis was conducted on the laboratory dataset. The goal of this analysis was twofold: first, to identify redundancy or complementarity among features, and second, to assess how individual features relate to near-term SOH decline.

The analysis was carried out on a per-cycle basis using all available cells in the lab dataset. For each battery, the SOH trajectory was smoothed using a Gaussian filter with standard deviation $\sigma = 1$ to reduce measurement noise and highlight underlying degradation trends. For each cycle, a fixed-length segment of 20 cycles was considered, and the slope of the SOH trajectory across this window was computed as:

$$\Delta\text{SOH} = \frac{\text{SOH}_{t+20} - \text{SOH}_t}{20} \quad (3.1)$$

This SOH slope was used as a proxy for short-term degradation rate. Simultaneously, the corresponding operational features, mean and standard deviation of voltage, current, and temperature, were extracted from each cycle to construct a dataset of feature–response pairs.

To evaluate statistical dependencies between features and with the degradation slope, Spearman’s rank correlation coefficient was used. Spearman correlation assesses monotonic relationships between variables without assuming linearity, making it suitable for operational battery data that may exhibit nonlinear trends. The final correlation matrix was computed across all cells and cycles in the dataset, producing a global view of feature co-dependence and relevance. Mathematically, Spearman’s correlation coefficient ρ between two variables X and Y is defined as the Pearson correlation between their ranked values:

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \quad (3.2)$$

where d_i is the difference between the ranks of X and Y for observation i , and n is the number of observations.

3.3 Model approaches and architectures

To evaluate the effectiveness of different neural network configurations for battery health forecasting, several Long Short-Term Memory (LSTM)-based architectures were designed and assessed, as illustrated in Figure 3.2. These models aim to capture both temporal degradation patterns and the influence of operational conditions on battery State of Health (SOH). The goal for these models are simply to predict the SOH for the next cycle, where the RUL forecasting is developed through an iterating sliding window approach, see section 3.5.

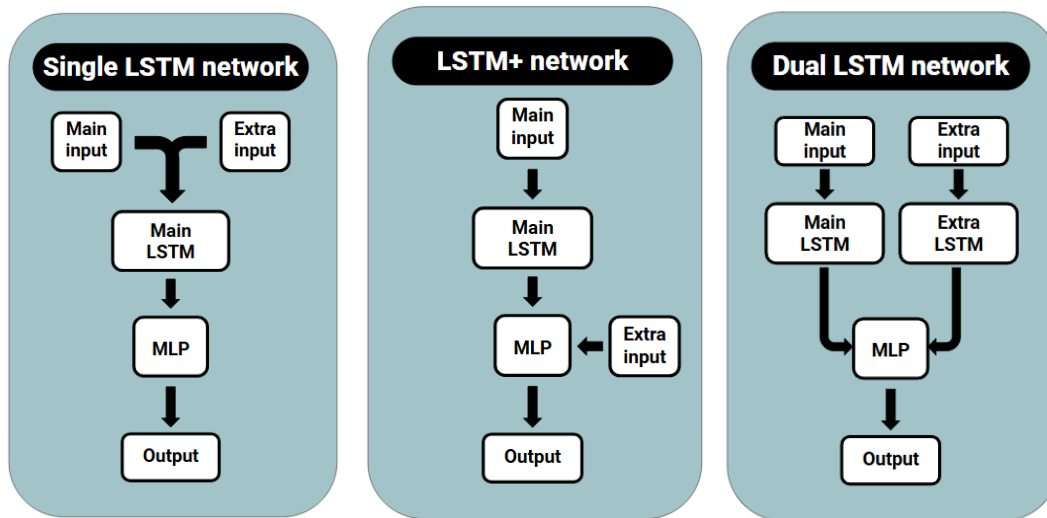


Figure 3.2: Comparison of the different LSTM architectures which are examined. In this thesis the main input consists of State of health data and the extra inputs consists of the features discussed in 3.2.

The first architecture considered is a single-stream LSTM model. This model processes a time series composed of SOH values and the variables described in section 3.2. The sequence is passed through a two-layer LSTM network with a hidden size of 100 units per layer. The hidden state from the final time step is then fed into a fully connected layer to produce a prediction for the next SOH value. This serves as a baseline for the LSTM approach.

A second approach extends the first model by passing only the SOH data to the LSTM and passing the extra features and the output of the LSTM to a MLP that made the prediction. This model architecture was proposed in order to avoid unstable behavior observed in the first model.

Building on this, the primary model architecture proposed in this work employs a dual LSTM configuration. In this setup, two parallel LSTM branches are used, one dedicated to the SOH sequence, and the other to the extra features. Each stream consists of two LSTM layers with a hidden size of 100. The hidden states from the final time steps of both streams are concatenated and passed through a two-layer multilayer perceptron (MLP), which generates the final SOH prediction. This

dual structure enables the model to learn complementary temporal representations from both historical degradation data and the vehicle usage pattern. Similar model architectures have been proved successful with similar methodologies, where both LSTMs learn the dependencies of different feature sets and whose outputs are then combined; and where single model architectures have been proven 'overwhelmed' [25, 26, 27].

To improve generalization to EV conditions, a fine-tuned version of the dual-stream model was employed using transfer learning. After pretraining the model on laboratory datasets, the LSTM layers were frozen, and the MLP is fine-tuned using a small subset of early-life cycles from each EV battery. This procedure allows the model to retain general degradation knowledge while adapting to battery-specific behaviors with limited data. This method follows the same approach as the paper by Kim et al. [3].

3.4 Final Dual stream LSTM architecture

The proposed model employs a dual LSTM architecture to process the time-series inputs and fuse their learned representations for downstream predictions. As illustrated in Figure 3.2, the network consists of two parallel LSTM branches, one for the primary input sequence consisting of SOH and one for the features in table 3.2 followed by a small multi-layer perceptron (MLP) that produces the final prediction.

In the primary branch, the model receives an input tensor $\mathbf{x} \in \mathbb{R}^{B \times T \times D}$, where B denotes the batch size, T the sequence length, and D the feature dimensionality which is one for the primary branch since it only consists of SOH. This tensor is processed by an 2-layer LSTM, each layer possessing hidden size $H = 100$. The hidden and cell states $\mathbf{h}_0, \mathbf{c}_0 \in \mathbb{R}^{L \times B \times H}$ are initialized to zero. After propagating through the LSTM, the final time-step embedding is extracted as

$$\mathbf{h}^{\text{main}} = \text{LSTM}(\mathbf{x})_{[:, -1, :]} \in \mathbb{R}^{B \times H}.$$

Simultaneously, an auxiliary branch processes a secondary input tensor $\mathbf{e} \in \mathbb{R}^{B \times T' \times D'}$, where $D' = 6$ corresponds to six additional features. This branch comprises an 2-layer LSTM of hidden size $H' = 100$. Its initial states $\mathbf{h}'_0, \mathbf{c}'_0 \in \mathbb{R}^{L' \times B \times H'}$ are likewise zero-initialized. The embedding at the last time step is given by

$$\mathbf{h}^{\text{aux}} = \text{LSTM}_{\text{aux}}(\mathbf{e})_{[:, -1, :]} \in \mathbb{R}^{B \times H'}.$$

The two embeddings are concatenated along their feature dimensions to form

$$\mathbf{h}^{\text{cat}} = [\mathbf{h}^{\text{main}}; \mathbf{h}^{\text{aux}}] \in \mathbb{R}^{B \times (H + H')}.$$

This fused representation is then fed into a two-layer MLP. The first linear transformation maps \mathbf{h}^{cat} to an intermediate vector $\mathbf{z}_1 \in \mathbb{R}^{B \times M}$, where $M = 10$, via

$$\mathbf{z}_1 = \mathbf{W}_1 \mathbf{h}^{\text{cat}} + \mathbf{b}_1.$$

A LeakyReLU activation with negative slope $\alpha = 0.01$ is applied. The resulting activations are then projected to the 1-dimensional output space yielding a score vector $\in \mathbb{R}^{B \times C}$. The model was trained for 200 epochs with an initial learn rate of 0.001 that decreased with a factor of 0.9 every fourth epoch. A coarse hyperparameter sweep was performed to select hyperparameters.

3.5 Evaluation

The following section describes the evaluation methods, with which the models performances were derived. The main structure revolves around being able to draw conclusions regarding transfer learning performance. First the models were evaluated on a homogeneous test data set, to establish a baseline of performance. Then the models were trained on different cycling behaviors, and consequently evaluated on a previously unseen behavior. Lastly the finalized model was evaluated under varying retraining conditions on the EV-data.

3.5.1 Evaluation on lab data

To emulate the conditions of adapting to personal driving behaviors, it was decided to keep one of the batches from the lab data out of the training and use that batch as validation data. This means that the model has not been trained at all to the specific user behavior of the validation data. Batch 5 of the XJTU data set was chosen as validation data. Batch 5 was chosen in part due to its varying trajectories within the batch, but also due to it containing relatively few batteries, see figure 3.1. Batches with similar variability contain considerably more batteries, which would result in noticeably less training data if held out.

RUL forecasting was first performed on the baseline data set and then the validation set, using all models described in figure 3.2. Forecasting was performed by feeding data to the models from the previous 20 cycles. The predicted value was then appended to the history window and the oldest entry was dropped, and the process was repeated until the predicted SOH fell below 0.8 or one thousand forecast steps had been generated. For the baseline data this was performed for 41 prediction ‘start cycles’ between 20-70, and for the validation there was 51 prediction ‘start cycles’ between 80-150. These start values were derived based on the length of the total data, where too early and too late predictions were deemed redundant.

To account for cell-specific degradation trajectories, the pre-trained NN was fine-tuned for twenty epochs with data from the first 80 cycles. This was done by freezing the weights in the LSTM and only updating the weights in the MLP by minimizing RMSE error between predicted and true SOH. The same forecasting procedure was then applied to this fine-tuned model to produce a second prediction trajectory.

3.5.2 Evaluation on EV data

For the evaluation on EV data, the pretrained dual LSTM was fine-tuned on the first 40 cycles using two different approaches. The first approach was to fine-tune the model on all available cars, and the second were to fine-tune the model on a single car at a time. The main idea with the first approach was for the NN to be able to learn behavior between batteries and the idea with the second approach was to get a more personalized learning for each car. The predictions for the base model and the two fine-tuned models were then plotted together with the SOH for comparison.

4

Results and Discussion

Throughout the thesis process, several different architectures, models, and evaluation strategies have been developed and tested. This chapter presents a curated selection of the most successful configurations, along with their performance on both laboratory and real-world data. A key focus during model development has been achieving high plasticity; that is, ensuring that the model is capable of generalizing across a wide range of cycling behaviors and battery types. The models here are evaluated on entirely new operational regimes not seen during training, except for the baseline evaluation on the HKUST data set. This reflects the practical requirement for models to be able to adapt to diverse and previously unseen conditions when deployed in the field.

The results show that the proposed dual LSTM architecture consistently outperforms simpler models when confronted with varied degradation profiles. Additionally, the application of transfer learning using early-life data from target batteries to fine-tune a pre-trained model, yields notable improvements in both lab and real-world scenarios. These findings support the approach of building generalizable models that can be efficiently adapted to specific battery contexts with limited additional data.

4.1 Feature analysis

To assess the relationships between input features and battery degradation, Spearman rank correlations were computed separately for the MIT and XTJU laboratory datasets. Figure 4.1 presents the resulting correlation matrices, showing both inter-feature dependencies and correlations with short-term SOH slope, see section 3.2.

While both datasets exhibited similar individual feature correlations with degradation, their internal feature dependencies differed. These observations highlight that among battery data sets, differences in experimental design can lead to meaningful variation in feature behavior. Despite this, the most degradation relevant features remained consistent across both sets, supporting their inclusion in the modeling.

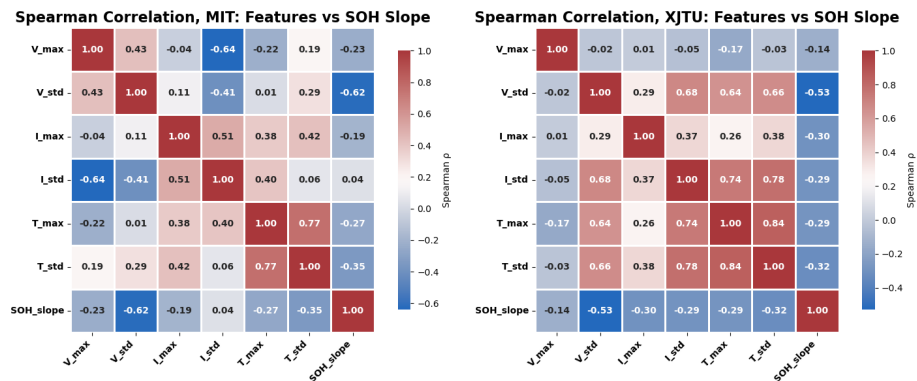


Figure 4.1: Spearman correlations between features for both lab-data sets.

4.2 Model performance on lab data

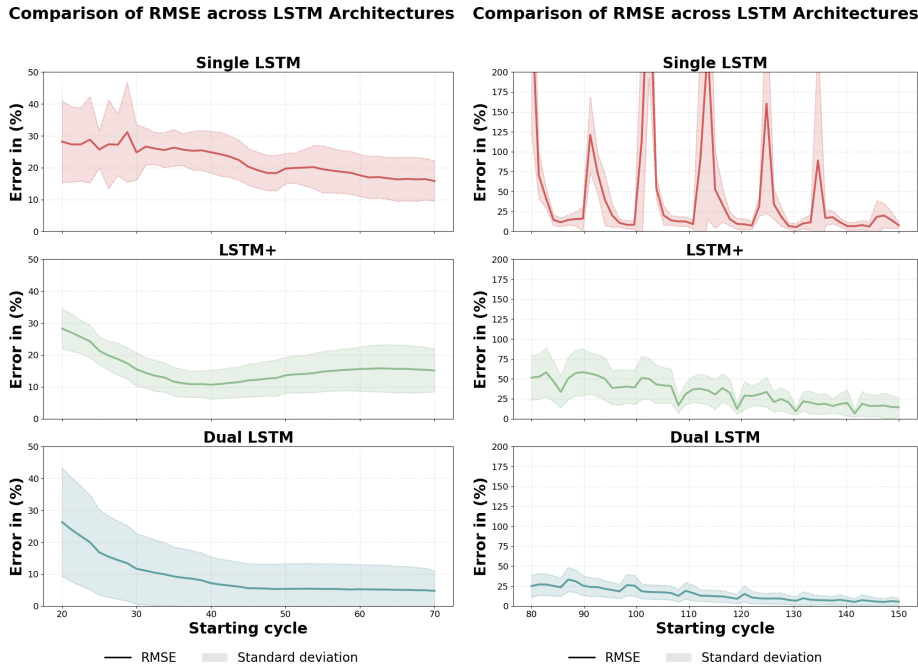
To get a baseline for the transfer learning evaluations, the model was first tested on a trial data set, shown in figure 4.2a. These results show that all models can learn RUL forecasting, with varying degree of uncertainty and error. The results convey that under forgiving conditions, RUL forecasting becomes a more or less trivial problem and all the tested models are moderately successful. Here the dual LSTM outperforms the other simpler models.

With a baseline established the models were evaluated with transfer learning in mind. The results from the three different architectures, described in 3.3, applied on the lab data is shown in figure 4.2b. Here the y-axis represents the average error in the form of a mean average percentage error (MAPE) between the predicted RUL and the true RUL. The x-axis shows from which cycle the model started its predictions. The shaded regions shows the standard deviation. The evaluation was conducted in accordance with the method described in section 3.5.

The results shows that the dual LSTM model also outperforms the two other architectures when presented with never before seen data and under similar training conditions. Some of the performance boost comes from lower MAPE score over all, but also considerably higher stability. One of the main challenges during forecasting on the batteries were diverging behavior, as seen in the single LSTM model; something which was overcome with the dual LSTM architecture.

When comparing the validation with the baseline test, there are some differences. All models show some unstable behavior in the new domain, while they were relatively stable in the baseline test. The measurement here is not purely the MAPE, but equally the stability of the forecasts.

When the dual LSTM model was finalized it was retrained and evaluated on the validation data once again. This comparison can be seen in figure 4.3a with the difference in 4.3b. From the data we can see that the retrained model performed 50%-20% better on its forecasts, with the largest performance boost coming from

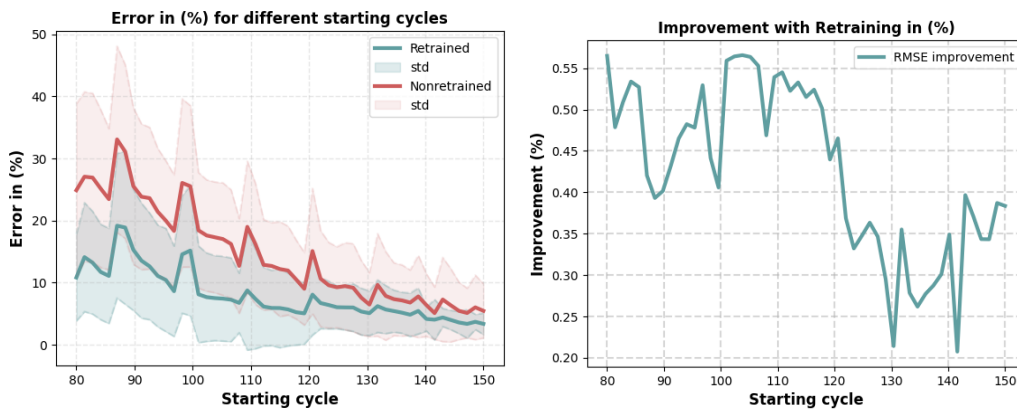


(a) RMSE scores for models evaluated on the baseline set. (b) RMSE scores for models evaluated on the validation set.

Figure 4.2: RMSE scores for the models when evaluated on different data sets. The plots to the left highlight the model performance when trained and validated on batteries with identical operating conditions. The plots to the right highlight the performance when transferred to a new domain.

the early lifetime forecasts. This validates the method of training a general model on several different usage and degradation behaviors and then being able to retrain to specific cases. There are still some unstable behavior which is not observed in the baseline tests, but to a lesser degree than the non retrained model.

Although comparisons are hard to draw to other papers, due to different data sets and validation methods, the paper by Kim et. al [3] is comparable. Their VarLSTM model was trained and retrained in a similar manner, although using different data sets. Their validation data is of comparable cycle length to the ones in this papers validation set, thus allowing for comparison. When they retrained on new data, their model got MAPE between 19.95-22.58% when presented with BOL data; while the dual LSTM model in this thesis got between 10-20% MAPE with BOL data. When confronted with MOL data, the VarLSTM got MAPE between 15.96-17.97%; and the dual LSTM got MAPE between 6-9%. Here BOL refers to the initial 30% of the cycle data and MOL refers to cycle data in the middle 40-60% of the total data. This shows that the proposed dual LSTM model shows increased performance when compared with contemporary models in the transfer learning regime.



(a) RMSE comparison between model and retrained model. (b) Improvement in RMSE by re-training.

Figure 4.3: Comparison of RMSE values and improvements for different starting cycles.

4.3 Transfer learning results on EV data

Figure 4.4 shows the dual-LSTM performance under different retraining conditions on the EV-data, see section 3.5.2. Since the data lacks EOL detail the performance here is more subjective, with no real evaluation metric. The retraining was done such that the model has only seen data points up until the start of the forecast, thus still ensuring validity to the results.

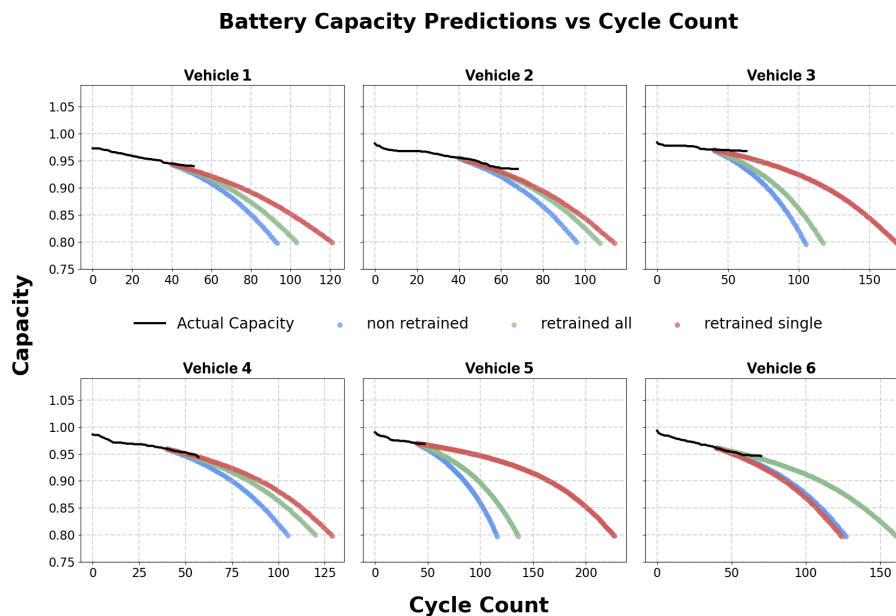


Figure 4.4: Model evaluation of the different retraining modes on the EV-data for the final dual-LSTM model.

5

Conclusion

This thesis set out to address a central challenge in battery health forecasting for EVs; namely the difficulty of training accurate and generalizable models when labeled real-world data is scarce and operational conditions are highly variable. To this end, the study explored how models trained on different degradation domains can adapt to previously unseen operating conditions using transfer learning.

A dual LSTM architecture was employed, capable of integrating both SOH trajectories and auxiliary operational features such as temperature, current, and voltage statistics. This model consistently outperformed simpler baselines, demonstrating robustness across differing battery types and cycling behaviors in lab settings. One of the key findings was the model’s ability to generalize to unseen operating regimes, an ability often lacking in existing literature where models are typically evaluated on narrow and homogeneous conditions.

Fine-tuning the pretrained models using a small amount of early-life data further improved accuracy, particularly in early-cycle forecasts, demonstrating a reduction in RMSE by up to 50%. This result highlights the value of lightweight personalization and supports the concept of deploying a global model across an EV fleet, with targeted updates to capture cell-specific behavior. Furthermore the proposed model seemingly outperforms at least one contemporary model in similar testing domains.

The chosen feature design, based on simple, statistical descriptors available from standard telemetry, proved sufficient to enable predictive performance. By avoiding reliance on features like incremental capacity or dQ/dV curves, which are common in academic studies but difficult to extract from field data, the model remains practical for real-world deployment.

Evaluation on real-world EV data showed promising results, with both collective and per-vehicle fine-tuning strategies yielding useful predictions, even in the absence of clearly defined end-of-life markers. However, the lack of complete degradation trajectories in the EV dataset limits how definitively model performance can be assessed. Future work would benefit from access to richer real-world datasets, particularly those including full cycle life data and ground-truth failure points.

Several directions remain open for future exploration. These includes exploring hy-

brid architectures such as CNN-LSTM or attention-based models, and expanding the transfer learning pipeline to support federated learning settings. Furthermore, research into features which are extractable in both laboratory and real-world domains, and which are more informative than the ones presented here, is needed to further enhance model performance.

In summary, this thesis demonstrates that a combination of generalizable architectures, replicable features, and targeted transfer learning can enable practical, scalable, and personalized SOH forecasting. This contributes a step toward reliable, fleet-level battery management systems that operate effectively despite data scarcity and operational variability.

Bibliography

- [1] M. A. Danzer, V. Liebau, and F. Maglia, “Aging of lithium-ion batteries for electric vehicles,” in *Advances in Battery Technologies for Electric Vehicles*, ch. 14, pp. 360–370, Elsevier Ltd., 2015.
- [2] C. A. Rufino Júnior, E. R. Sanseverino, P. Gallo, M. M. Amaral, D. Koch, Y. Kota, H.-G. Schweiger, and H. Zanin, “A comprehensive review of ev lithium-ion battery degradation,” *Preprints*, 2023.
- [3] S. Kim, Y. Y. Choi, K. J. Kim, and J.-I. Choi, “Forecasting state-of-health of lithium-ion batteries using variational long short-term memory with transfer learning,” *Journal of Energy Storage*, vol. 41, p. 102893, 2021.
- [4] F. von Bülow, *A Data-Driven Fleet Service: State of Health Forecasting of Lithium-Ion Batteries*. PhD thesis, University of Wuppertal, 2023.
- [5] H. Xu, L. Wu, and S. e. a. Xiong, “An improved cnn-lstm model-based state-of-health estimation approach for lithium-ion batteries,” *Energy*, vol. 276, p. 127585, 2023.
- [6] Y. Zhang, T. Wik, J. Bergström, and C. Zou, “Practical battery state of health estimation using data-driven multi-model fusion,” *IFAC-PapersOnline*, vol. 56, no. 2, pp. 3776–3781, 2023.
- [7] V. Steininger, K. Rumpf, P. Hüsson, W. Li, and D. U. Sauer, “Automated feature extraction to integrate field and laboratory data for aging diagnosis of automotive lithium-ion batteries,” *Cell Reports Physical Science*, vol. 4, p. 101596, 2023.
- [8] K. Liu, Q. Peng, Y. Che, Y. Zheng, K. Li, R. Teodorescu, D. Widanage, and A. Barai, “Transfer learning for battery smarter state estimation and ageing prognostics: Recent progress, challenges, and prospects,” *Advances in Applied Energy*, vol. 9, p. 100117, 2023.
- [9] K. A. Severson, P. M. Attia, and N. e. a. Jin, “Data-driven prediction of battery cycle life before capacity degradation,” *Nature Energy*, vol. 4, no. 5, pp. 383–391, 2019.

- [10] M. Ghaznavi, M. Alahmad, and Y. Chen, “State of health prognostics for series battery packs: A universal deep learning method,” *Journal of Energy Storage*, vol. 66, p. 107329, 2023.
- [11] F. Wang, Z. Zhai, Z. Zhao, Y. Di, and X. Chen, “Physics-informed neural network for lithium-ion battery degradation stable modeling and prognosis,” *Nature Communications*, vol. 15, no. 1, p. 4332, 2024.
- [12] X. Li, D. Yu, V. S. Byg, and S. D. Ioan, “The development of machine learning-based remaining useful life prediction for lithium-ion batteries,” *Journal of Energy Chemistry*, vol. 82, pp. 103–121, 2023.
- [13] G. Cheng, X. Wang, and Y. He, “Remaining useful life and state of health prediction for lithium batteries based on empirical mode decomposition and a long and short memory neural network,” *Energy*, vol. 232, p. 121022, 2021.
- [14] Y. Ji, Z. Chen, Y. Shen, K. Yang, Y. Wang, and J. Cui, “An rul prediction approach for lithium-ion battery based on sade-mesn,” *Applied Soft Computing*, vol. 104, p. 107195, 2021.
- [15] L. Shen, J. Li, L. Meng, L. Zhu, and H. T. Shen, “Transfer learning-based state of charge and state of health estimation for li-ion batteries: A review,” *IEEE Transactions on Transportation Electrification*, vol. 10, no. 1, pp. 1465–1481, 2023.
- [16] G. Ma, S. Xu, T. Yang, Z. Du, L. Zhu, H. Ding, and Y. Yuan, “A transfer learning-based method for personalized state of health estimation of lithium-ion batteries,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 1, pp. 759–769, 2022.
- [17] D. Pan, H. Li, and S. Wang, “Transfer learning-based hybrid remaining useful life prediction for lithium-ion batteries under different stresses,” *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–10, 2022.
- [18] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [19] L. Yao, J. Wen, S. Xu, J. Zheng, J. Hou, Z. Fang, and Y. Xiao, “State of health estimation based on the long short-term memory network using incremental capacity and transfer learning,” *Sensors*, vol. 22, p. 7835, 2022.
- [20] I. Sutskever, O. Vinyals, and Q. V. Le, “Sequence to sequence learning with neural networks,” in *Advances in Neural Information Processing Systems (NeurIPS)*, pp. 3104–3112, 2014.
- [21] T. Fischer and C. Krauss, “Deep learning with long short-term memory networks for financial market predictions,” *European Journal of Operational Research*, vol. 270, no. 2, pp. 654–669, 2018.

- [22] A. Zhang, Z. C. Lipton, M. Li, and A. J. Smola, “Dive into deep learning,” 2024.
- [23] F. Wang, J. Zhang, and Z. Zhou, “State of health estimation based on the long short-term memory network using incremental capacity and transfer learning,” *Journal of Energy Storage*, vol. 55, p. 105134, 2023.
- [24] X. Tang, X. Lai, C. Zou, Y. Zhou, J. Zhu, Y. Zheng, and F. Gao, “Detecting abnormality of battery lifetime from first-cycle data using few-shot learning,” *Advanced Science*, vol. 11, no. 6, p. 2305315, 2024.
- [25] Z. Shi and A. Chehade, “A dual-lstm framework combining change point detection and remaining useful life prediction,” *Reliability Engineering & System Safety*, vol. 205, p. 107257, 2021.
- [26] H. Alharkan, S. Habib, and M. Islam, “Solar power prediction using dual stream cnn-lstm architecture,” *Sensors*, vol. 23, no. 2, p. 945, 2023.
- [27] R. Jin, Z. Chen, K. Wu, M. Wu, X. Li, and R. Yan, “Bi-lstm-based two-stream network for machine remaining useful life prediction,” *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–10, 2022.

DEPARTMENT OF SOME SUBJECT OR TECHNOLOGY
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden
www.chalmers.se



CHALMERS
UNIVERSITY OF TECHNOLOGY