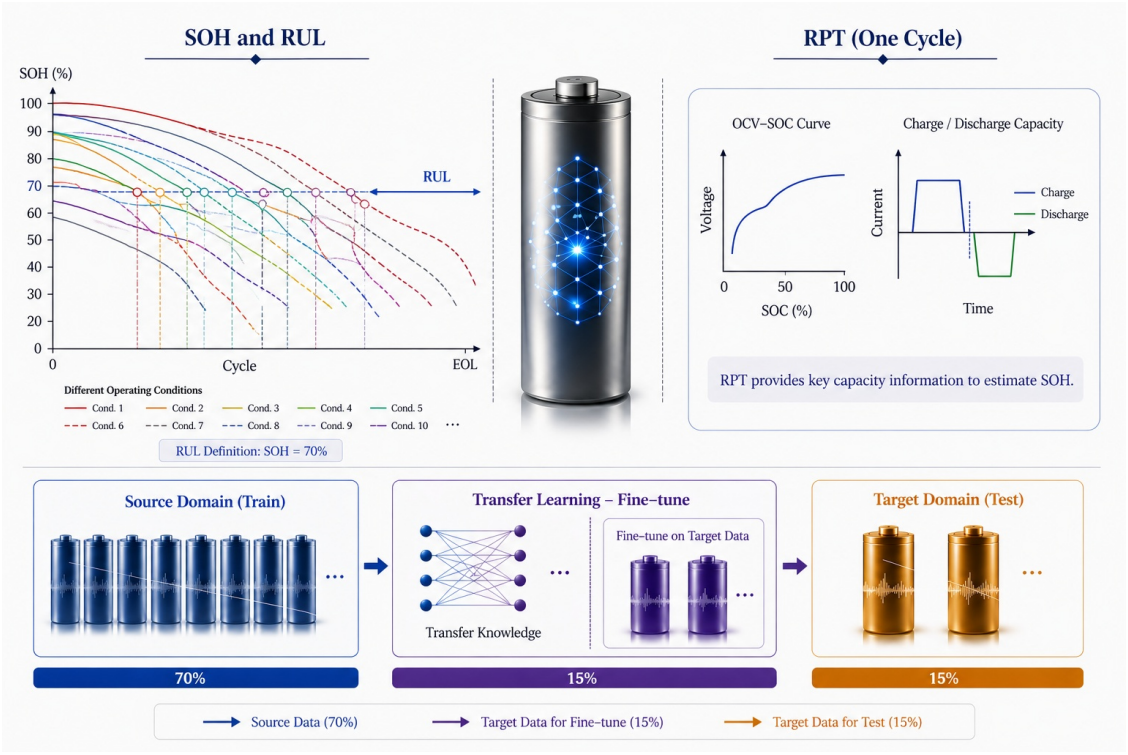
CHALMERS
UNIVERSITY OF TECHNOLOGY



Transfer Learning for Battery Health Prediction under Varying Operating Conditions

Master's thesis in Mobility Engineering

Chunqiu Xia

DEPARTMENT OF Electrical Engineering
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2026
www.chalmers.se

Master's thesis 2026

Transfer Learning for Battery Health Prediction under Varying Operating Conditions

Chunqiu Xia

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Electrical Engineering
Chalmers University of Technology
Gothenburg, Sweden 2026

Transfer Learning for Battery Health Prediction
under Varying Operating Conditions
Chunqiu Xia

Supervisor: Qingbo Zhu, Department of Electrical Engineering; Shengyu Tao, Department of Electrical Engineering
Examiner: Changfu Zou, Department of Electrical Engineering

Master's Thesis 2026
Department of Electrical Engineering
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Cover: Conceptual visualization of battery state of health (SOH) estimation and remaining useful life (RUL) prediction under cross-condition settings. The cover illustrates heterogeneous SOH degradation trajectories, the 70% SOH-based RUL definition, RPT-based capacity/SOH characterization, and a transfer learning workflow from source domain training to target domain fine-tuning and testing.

Typeset in LaTeX
Printed by Chalmers Reproservice
Gothenburg, Sweden 2026

Transfer Learning for Battery Health Prediction under Varying Operating Conditions

Chunqiu Xia
Department of Electrical Engineering
Chalmers University of Technology

Abstract

Battery health estimation and lifetime prediction are essential for the safe and reliable operation of Li-ion batteries. In practical applications, predictive models are often required to generalize across different cells or cell groups, where operating conditions, materials, and degradation behaviours may vary. This thesis investigates battery health modelling from a cross-condition perspective under constant-current operating protocols, where charge C-rate, discharge C-rate, and depth of discharge define the domain structure.

The study first examines state of health (SOH) estimation using features extracted from incremental capacity (IC) curves. Although SOH estimation provides useful diagnostic information, it shows limited domain divergence in the present dataset, mainly because most available samples are concentrated in the early ageing stage where degradation trajectories across-conditions remain similar. Therefore, while still considering SOH estimation as an important diagnostic task, this thesis also considers remaining useful life (RUL) prediction because it can construct sufficiently large domain divergence to support the transfer learning objective.

A transfer learning framework is then developed for cross-condition RUL prediction. The model is first trained on a source cell group and subsequently adapted to a target group through fine-tuning. Optuna with the Tree-structured Parzen Estimator (TPE) is employed to optimize model structure and learning hyperparameters, and a sensitivity analysis is conducted to examine the influence of early-stage data availability on prediction performance.

The results show that direct cross-condition prediction is challenging and that models trained without transfer learning have limited generalization capability. Fine-tuning improves RUL prediction on the target group, demonstrating the practical value of transfer learning for cross-condition battery prognostics. The results also indicate that prediction performance is affected by feature selection, domain definition, and the quantity and selection of cells used for fine-tuning.

Keywords: Li-ion batteries, battery health prediction, state of health, remaining useful life, transfer learning, hyperparameter optimization, operating conditions

Acknowledgements

I would like to express my sincere gratitude to my supervisors and examiner for their guidance, advice, and generous support throughout this thesis. In particular, I would like to thank my examiner, Changfu Zou, for his continuous support and dedication to my thesis work. From him, I received not only the initial thesis idea and valuable technical guidance, but also important lessons about research attitude and academic qualities. I am also deeply thankful to Qingbo Zhu and Shengyu Tao, who have supported me throughout this process. They helped me not only complete the thesis work, but also learn how to present and express research ideas in an academic style. These skills will be extremely valuable for my future PhD studies and research career. In addition, I would like to thank Albert Škegro and Francesco Popolizio for sharing their knowledge in group meetings and for warmly welcoming me when I joined the team, which gave me a strong sense of belonging during this period.

I also want to thank my family for supporting me in coming here, as the Nordic countries have always been a dream place for them to live. At the same time, I could not have adapted to this new environment without my friends. Their encouragement and companionship helped me continue pursuing my research dream. A special thank you goes to Keran Hu and Xinlei Chen. Thank you for believing in me and for being part of this path.

Chunqiu Xia, Gothenburg, 06/2026

Acronyms

CC	Constant current
CV	Constant voltage
DoD	Depth of discharge
DV	Differential voltage
DVA	Differential voltage analysis
EFC	Equivalent full cycle
EOL	End-of-life
HPO	Hyperparameter optimization
IC	Incremental capacity
ICA	Incremental capacity analysis
ISU-ILCC	Iowa State University-Iowa Lakes Community College
Li-ion	Lithium-ion
MAE	Mean absolute error
MAPE	Mean absolute percentage error
ML	Machine learning
MLP	Multilayer perceptron
NMC/Gr	Nickel manganese cobalt oxide/graphite
RPT	Reference performance test
RUL	Remaining useful life
sMAPE	Symmetric mean absolute percentage error
SOH	State of health
TPE	Tree-structured Parzen Estimator
wMAPE	Weighted mean absolute percentage error

Nomenclature

Indices

i	Index for prediction samples or observations
k	Index for an observation point along a battery ageing trajectory
n	Number of samples used in a metric calculation
t	Time index, expressed in weeks unless otherwise stated
$w0$, $w3$	Week 0 and week 3 reference points used in early-life feature

Sets and Domains

D	Generic machine learning domain
$\mathcal{D}_s$	Source domain used for pretraining or source domain modelling
$\mathcal{D}_t$	Target domain used for fine-tuning and target-test evaluation
D_{train}	Training dataset used during hyperparameter optimization
D_{val}	Validation dataset used during hyperparameter optimization
$\mathcal{X}$	Feature space
$\mathcal{Y}$	Label or target space
Λ	Hyperparameter search space

Battery and Feature Quantities

C_0	Reference capacity of a fresh cell
C_k	Capacity measured at observation point k
C_{chg}	Charging C-rate used as a cycling condition feature
DoD	Depth of discharge
Q	Discharge capacity
$Q_{\text{interpolated}}$	Interpolated discharge capacity measurement used for feature extraction
V	Cell voltage
$\text{IC}(V)$	Incremental capacity response as a function of voltage
$\text{d}Q/\text{d}V$	Incremental capacity derivative
$\Delta dQ/dV$	Change in the incremental capacity response between selected reference weeks
ΔQ	DVA-based capacity change descriptor
$\Delta\text{CV Time}$	Change in constant voltage hold time
$f1$–$f10$	Engineered feature family considered in the modelling workflow

$\mathbf{x}_k$	Input feature vector at observation point k
y_k^{SOH}	state of health (SOH) target label at observation point k
y_k^{RUL}	remaining useful life (RUL) target label at observation point k
t_k	Current observation time or week
t_{EOL}	Time or week at which the end-of-life (EOL) threshold is reached

Model and Optimization Quantities

A_λ	Learning algorithm configured by hyperparameter vector λ
$f(\cdot)$	Predictive function mapping input features to target values
λ	Hyperparameter configuration
λ^*	Optimal hyperparameter configuration selected
λ_{next}	Next hyperparameter configuration proposed by the TPE
$\mathcal{L}_{\mathrm{val}}$	Validation loss used as the hyperparameter optimization
y	Objective value observed from one hyperparameter Optimization trial
y^*	TPE threshold separating good trials from the remaining trials
$p(\lambda \mid y)$	Conditional density of hyperparameters given the objective value
$l(\lambda)$	Density estimated from good TPE trials
$g(\lambda)$	Density estimated from the remaining TPE trials
γ	Quantile probability defining the TPE threshold, $p(y < y^*)$
$\mathrm{EI}_{y^*}(\lambda)$	Expected improvement for threshold y^*

Evaluation Metrics

y_i	True target value of sample i
$\hat{y}_i$	Predicted target value of sample i
MAE	Mean absolute error
MAPE	Mean absolute percentage error
sMAPE	Symmetric mean absolute percentage error
wMAPE	Weighted mean absolute percentage error

Contents

1

Introduction

1.1 Background

Lithium-ion batteries have become a core energy storage technology in electrical mobility, portable electronics, and many other applications that require high energy density and reliable long term operation. As battery systems are used over extended periods, however, their capacity gradually fades and their internal behaviour changes. These degradation processes affect both safety and functional availability, making battery health monitoring an essential part of battery management and lifetime assessment [1, 2, 3].

Two closely related prediction tasks are particularly important in this context. State of health (SOH) estimation describes the current degradation state of a battery, often through a capacity-based indicator that summarizes how far the cell has aged relative to its fresh condition. Remaining useful life (RUL) prediction goes one step further by estimating how much usable life remains before a chosen end-of-life (EOL) threshold is reached. Together, these tasks support both diagnostics and prognostics: SOH estimation helps characterize the present condition of a cell, whereas RUL prediction helps anticipate its future degradation trajectory.

In data-driven battery research, the quality of prediction depends strongly on the relationship between the available measurements and the underlying degradation behaviour. Under controlled constant-current cycling protocols, repeated reference measurements and voltage and capacity signals provide an opportunity to engineer informative health-related features. At the same time, practical deployment rarely occurs under perfectly homogeneous conditions. Batteries may differ in charge C-rate, discharge C-rate, depth of discharge (DoD), degradation histories, or cell group characteristics, while predictive models are still expected to remain informative across such variation.

This challenge motivates the present thesis. Rather than studying battery health prediction only within a single homogeneous dataset, the work investigates cross-condition modelling under heterogeneous constant-current operating conditions. In this setting, charge C-rate, discharge C-rate, and DoD form the axes used to separate battery groups into related but distinct domains. The central issue is therefore not only whether battery health indicators can be predicted accurately, but also whether predictive knowledge learned from one group of cells can be transferred usefully to another.

1.2 Motivation

The thesis is motivated by an empirical and methodological observation. On the one hand, SOH estimation is a natural and widely used as starting point for battery health modelling, because it provides a direct and interpretable diagnostic target. On the other hand, preliminary experiments in the present dataset suggest that SOH is not always the most informative target for studying cross-condition transferability.

In the considered battery groups, much of the available data are concentrated in the early-stage of ageing, where the degradation trajectories of different groups remain relatively similar. Under such conditions, a model may still achieve reasonable SOH estimates without clearly exposing the domain divergence that later becomes important for cross-condition generalization. As a result, the SOH task can provide useful diagnostic insight while still being too mild to reveal the full difficulty of source to target transfer.

This observation motivates a shift in emphasis from SOH estimation to RUL prediction. Compared with SOH, RUL prediction is more sensitive to future degradation behaviour and therefore provides a stricter test of whether the source and target domains differ in a way that matters for prognosis. It also makes the transfer learning problem more relevant from an application perspective, since practical battery management decisions often depend on how much useful lifetime remains rather than only on the current health level.

A second source of motivation concerns limited target domain data. In realistic transfer settings, labelled data from the target group are often scarce, and this scarcity makes it difficult to train robust models from scratch. Fine-tuning-based transfer learning therefore becomes attractive as a way to reuse source domain knowledge while still adapting to target specific behaviour. The thesis is thus motivated not only by the need for accurate prediction, but also by the need for a practical modelling strategy under heterogeneous domains and constrained target domain information.

1.3 Research Objective

The overall objective of this thesis is to develop and evaluate a data-driven framework for cross-condition battery prognostics under heterogeneous constant-current operating conditions. The framework is designed to move from an initial diagnostic SOH study to a final prognostic RUL study in order to determine which target definition is more informative for analyzing domain transfer.

More specifically, the thesis aims to: first, examine whether engineered features extracted from battery test data provide meaningful information for SOH estimation and RUL prediction; second, evaluate how difficult direct cross-condition prediction is when no transfer mechanism is used; and third, assess whether a fine-tuning-based transfer learning strategy can improve target domain RUL prediction when only limited target domain data are available.

A further objective is to make the modelling pipeline systematic rather than purely heuristic. For this reason, hyperparameter optimization with Optuna us-

ing the Tree-structured Parzen Estimator (TPE) sampler is incorporated into the workflow, and sensitivity analyses are used to examine how prediction performance changes with data availability and stochastic variation. Taken together, these objectives define the thesis as a study of both predictive performance and transferability across battery groups.

1.4 Research Questions

- *RQ1:* How does battery health prediction generalize across different operating conditions?
- *RQ2:* Is RUL prediction more suitable than SOH estimation for studying cross-condition domain shift?
- *RQ3:* Can fine-tuning-based transfer learning improve cross-condition RUL prediction with limited target condition data?

1.5 Scope and Limitations

The scope of this thesis is limited to controlled battery cycling data collected under predefined constant-current operating protocols. Within this setting, the domain split is based on the operating condition axes of charge C-rate, discharge C-rate, and DoD, rather than on a test environment distinction. The study focuses on cross-condition modelling within the available dataset and therefore does not claim direct generalization to arbitrary field conditions, dynamic drive cycles, or battery chemistries outside those represented in the data.

Therefore, the present conclusions should be interpreted within the constant-current ageing domain. Extension to realistic EV dynamic profiles requires additional validation because dynamic load shape, rest periods, regenerative pulses, and low-frequency current components may alter degradation trends [4, 5].

The work is also limited by the way domains are defined. In this thesis, domains are constructed from battery groups and operating condition partitions available in the dataset. This is a reasonable and practically useful definition for the present study, but it is not the only possible one. Alternative domain definitions could lead to somewhat different transfer learning conclusions.

In addition, the thesis concentrates on data-driven regression using engineered battery features and a modelling pipeline based on neural network. The goal is not to exhaustively compare all possible machine learning models, nor to claim that the adopted feature set is universally optimal. Instead, the selected workflow is used as a consistent experimental framework for investigating the transition from SOH estimation to RUL prediction and for evaluating the role of fine-tuning-based transfer learning.

Finally, the study is constrained by target domain data availability. Since only limited target domain data are assumed to be available for adaptation, the reported results should be interpreted as evidence for the usefulness of transfer learning under

data scarcity rather than as an upper bound on the achievable performance with abundant target domain supervision.

1.6 Literature Review

1.6.1 Battery Diagnostics and Prognostics under Controlled Cycling Protocols

Recent review studies consistently describe battery health prediction as a combination of diagnostic and prognostic tasks. In this literature, state of health (SOH) estimation is commonly treated as a present-state diagnostic problem, whereas remaining useful life (RUL) prediction is treated as a forward-looking prognostic problem that depends on the future evolution of degradation toward a predefined end-of-life (EOL) threshold [6, 7, 8, 9, 10]. This distinction is important because the two tasks are related but not identical. A battery can exhibit a certain current health level while still following a future degradation trajectory that differs substantially from that of another battery with a similar present SOH. For this reason, SOH and RUL are generally regarded as complementary indicators rather than interchangeable prediction targets.

Under controlled cycling protocols, the difference between diagnostics and prognostics becomes especially clear. Controlled experiments make it possible to obtain repeated capacity measurements, voltage-capacity trajectories, and degradation records under relatively stable testing conditions. Such data are useful for both SOH estimation and RUL prediction, but the modelling emphasis differs. SOH estimation mainly focuses on mapping current measurements or extracted health indicators to an interpretable health index, whereas RUL prediction must infer the remaining trajectory length before EOL is reached. Therefore, RUL prediction is more sensitive to nonlinear ageing behaviour, uncertainty accumulation, and cell-to-cell variability [11, 12, 9, 13, 14].

This distinction is particularly relevant for cross-condition battery health modelling. Two cell groups may appear similar from a pointwise SOH perspective while still differing materially in how quickly they approach EOL. In early-life data, degradation trajectories may not yet be sufficiently separated, making SOH estimation relatively insensitive to some domain differences. RUL prediction, by contrast, depends directly on the future degradation path and is therefore often a stricter task when the goal is to understand how batteries diverge under different operating conditions or group structures [7, 15, 16, 17, 18]. This motivates the present thesis to consider both SOH estimation and RUL prediction, while placing greater interpretive weight on RUL when evaluating cross-condition generalization.

1.6.2 Feature Engineering for Battery State of Health Estimation and Remaining Useful Life Prediction

A central finding in the battery health literature is that predictive performance depends strongly on feature representation. Although raw current, voltage, temper-

ature, and time-series data can be learned directly by advanced models, many studies still rely on engineered health indicators because they summarize degradation-related information in a more structured and interpretable form [19, 20, 21, 22, 9, 10]. Common feature families include capacity fade measures, internal resistance indicators, time-based statistics from charge or discharge curves, and derivative features extracted from voltage-capacity relations.

Among these feature families, incremental capacity analysis (ICA, $\mathrm{d}Q/\mathrm{d}V$) and differential voltage analysis (DVA, $\mathrm{d}V/\mathrm{d}Q$) remain particularly influential. Prior work shows that peak locations, peak magnitudes, and shape changes in IC or DV curves can reflect ageing-induced electrochemical changes and can therefore be used for SOH estimation, capacity estimation, and related prognostic tasks [23, 24, 25]. Such features are especially attractive in controlled cycling datasets because repeated reference performance tests provide relatively clean voltage-capacity curves from which derivative-based features can be extracted. In addition, recent work on battery SOH transfer learning continues to combine voltage-derived health indicators with machine learning models, suggesting that physically interpretable feature design remains valuable even when neural networks are used [26].

At the same time, the literature also shows that ICA- and DVA-type features are condition-dependent rather than universally stable. Their quality can degrade when the current rate is high, the signal is noisy, the voltage window is only partially observed, or the ageing test does not provide sufficiently comparable diagnostic cycles [27, 24]. This limits straightforward deployment under arbitrary field conditions and makes controlled cycling a particularly suitable setting for derivative-curve-based feature engineering. In the context of this thesis, these findings support the use of health-related features extracted from structured constant-current cycling measurements. They also motivate careful feature selection, because the objective is not only to obtain good within-group prediction performance but also to evaluate whether the selected features remain informative when the model is applied across different battery groups [18].

1.6.3 Operating-Condition Heterogeneity and Domain Shift in Battery Ageing Data

Battery ageing is highly dependent on operating conditions. Factors such as charging C-rate, discharging C-rate, depth of discharge, temperature, voltage window, and rest conditions can affect the balance among degradation mechanisms and lead to different capacity-fade trajectories [12, 16, 18]. In many data-driven studies, these operating variables are included as input features so that a single model can learn their relationship with lifetime or SOH. This approach is useful when the training and test data cover similar regions of the operating-condition space. However, when a model is trained on one subset of cells and tested on another subset with different condition distributions, the problem becomes more than ordinary supervised prediction.

From a machine learning perspective, this situation can be interpreted as domain shift. The source and target domains may share the same prediction target, such as SOH or RUL, but differ in the distribution of input features, degradation trajec-

tories, or feature-target relationships. In battery ageing data, such a shift can arise because different cycling protocols do not merely change the numerical value of an input variable; they can also alter the ageing path itself. For example, two cells with similar early-cycle health indicators may still reach EOL at different times if their operating conditions induce different degradation rates. Therefore, good interpolation performance within a matched-condition dataset does not necessarily imply reliable generalization across operating-condition groups.

This issue is especially important for early-life battery prediction. Early ageing data are valuable because they enable earlier decision-making, but they may also provide limited information about later divergence among cells. If the source and target groups are separated by operating conditions, early-stage features may not fully capture how the groups will diverge over the rest of life. This makes cross-condition generalization a stricter and more practically relevant evaluation than random sample splitting. It also motivates experimental designs in which source domains, target fine-tuning data, and target test data are explicitly separated by operating-condition groups rather than mixed randomly.

For the present thesis, operating-condition heterogeneity is therefore treated not only as a feature-design problem but also as a domain-definition problem. The central question is whether a model trained on one battery group can be adapted to another group when only limited labelled target data are available. This framing connects battery ageing science with transfer learning: if operating conditions induce systematic differences between cell groups, then the reuse and adaptation of source-domain knowledge becomes a natural strategy for improving target-domain prediction.

1.6.4 Transfer Learning for Battery Health Diagnostics and Prognostics

Transfer learning has attracted growing attention in battery health research because labelled degradation data are expensive to collect and because models trained on one set of cells often generalize poorly to another. In the broader machine learning literature, transfer learning is commonly motivated by the reuse of source-domain knowledge when target-domain labels are limited, while domain-adaptation methods address distribution mismatch between source and target data [28, 29, 30]. In the battery state-estimation literature, recent transfer learning methods are commonly grouped into three broad categories: fine-tuning methods, metric-based methods, and adversarial adaptation methods [31, 32, 33]. Figure 1.1 adapts this taxonomy and highlights the route selected in the present thesis.

Fine-tuning methods first pretrain a model on a source domain and then adapt part or all of the model parameters using limited labelled target-domain data. This strategy is well matched to the setting of the present thesis because both the source-domain training set and the target-domain adaptation set are labelled. Under this condition, supervised parameter adaptation provides a direct and reproducible way to test whether source-domain knowledge can improve target-domain prediction. Compared with adversarial adaptation, fine-tuning does not require an additional discriminator or an explicit domain-confusion objective. Compared with metric-

based methods, it does not rely primarily on constructing a distance metric to align feature distributions. This makes the modelling pipeline easier to interpret and easier to evaluate in a controlled thesis setting [28, 30, 34, 26, 33].

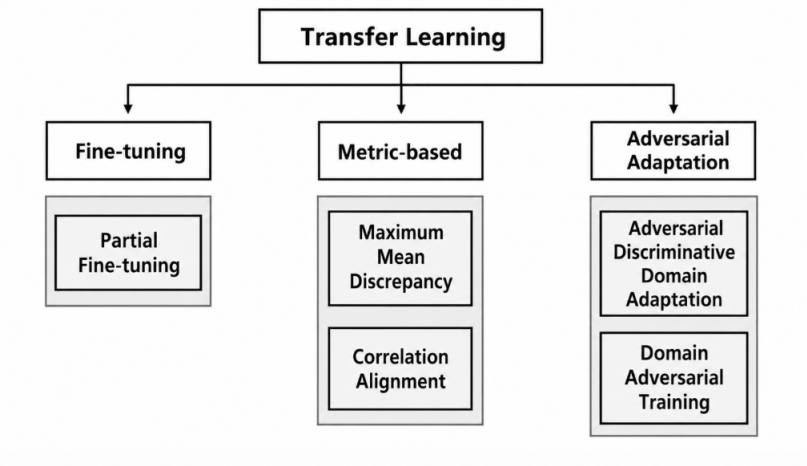


Figure 1.1: Adapted taxonomy of transfer learning strategies for battery state of health (SOH) and remaining useful life (RUL) estimation. Recent review work groups battery transfer learning methods into three broad categories: fine-tuning methods, metric-based methods, and adversarial adaptation methods. The present thesis follows the fine-tuning route because it best matches the labelled-data setting and the experimental objective of controlled cross-condition comparison [33].

Fine-tuning is also suitable for controlled comparison. The effect of transfer can be assessed directly by comparing a source-pretrained and target-fine-tuned model against benchmark models trained without the same pretraining procedure. This is important because the objective of the present thesis is not merely to obtain the lowest prediction error, but to determine whether transferred knowledge offers measurable benefit under cross-condition shift. From an application perspective, fine-tuning is also practical because a model pretrained on historical cell data could be adapted to a new cell group using only limited labelled target data, reducing the need for extensive additional ageing experiments [34, 35, 26].

This choice does not imply that metric-based or adversarial approaches are unimportant. These methods remain valuable alternatives for handling domain discrepancy and help clarify the broader methodological landscape of battery transfer learning [33]. However, for the present thesis, fine-tuning provides the clearest balance among methodological simplicity, labelled-data compatibility, practical deployability, and fair comparison against non-transfer benchmarks. For this reason, it is adopted as the main transfer learning mechanism in the reported cross-condition SOH and RUL study.

1.6.5 Model Selection and Sensitivity

Model selection is an important methodological issue in data-driven battery health modelling because prediction performance depends not only on the selected input

features and target definition, but also on model configuration. Learning rate, number of hidden layers, hidden dimensions, regularization strength, dropout, batch size, and training duration can all affect the final SOH or RUL prediction result. Manual tuning is difficult to justify in a transfer learning study, where benchmark and fine-tuned models must be compared under a consistent experimental design. A systematic search procedure is therefore needed so that performance differences can be attributed more fairly to the modelling strategy rather than to arbitrary parameter choices [36, 37, 38].

This thesis uses Optuna with the Tree-structured Parzen Estimator (TPE) sampler because this combination is well suited to the conditional structure of the model selection problem. In the present workflow, some hyperparameters are only meaningful under certain previous choices. For example, the number of layers determines how many hidden-layer widths are active, and fine-tuning-related settings become relevant only after a source model has been pretrained. Such a search space is naturally tree-structured rather than purely rectangular. TPE is designed for this kind of conditional hyperparameter space, and Optuna provides a convenient define-by-run framework for implementing conditional search logic [39, 38]. For this reason, Optuna with the TPE sampler is adopted as a practical and reproducible way to tune both model architecture and training behaviour in the later RUL transfer learning experiments.

Sensitivity analysis is also needed because battery prognostics depends strongly on data availability. A model that performs well when many observations are available may not remain reliable when only early-life information is used. This issue is especially important for transfer learning: if target-domain data are limited, the benefit of transferring source-domain knowledge may depend on the observation week, the number of target cells used for adaptation, and the particular target cells selected for fine-tuning. Therefore, the literature motivates not only systematic hyperparameter optimization, but also sensitivity analyses on early-week information and fine-tuning data selection. These analyses are necessary for evaluating whether the observed transfer benefit is robust or only a consequence of a favourable data split.

1.6.6 Research Gap Statement

The reviewed literature shows that SOH estimation and RUL prediction have been widely studied, and recent work has demonstrated that early ageing data can be used to predict battery lifetime under varying usage conditions [18]. However, many existing studies still evaluate models under relatively matched conditions or treat operating-condition variables mainly as ordinary input features. Less attention has been given to the question of whether operating-condition differences themselves should be treated as a source of domain shift for SOH estimation and RUL prediction. In other words, the effects of charging C-rate, discharging C-rate, and depth of discharge are often considered in prediction, but they are not always studied explicitly as a cross-condition transfer problem.

Building on the dataset and motivation of *Predicting battery lifetime under varying usage conditions from early aging data*, this thesis introduces transfer learning to

examine that gap more directly [18]. The key question is whether a model trained on one operating-condition group can be adapted to another group using limited labelled target-domain data. Rather than assuming that a model trained under one condition will generalize automatically, the thesis evaluates whether source-domain knowledge can be reused through fine-tuning to improve target-domain SOH and RUL prediction.

The research gap can therefore be summarized as follows: existing battery health studies provide strong foundations for feature-based SOH estimation, RUL prediction, and lifetime modelling under varying usage conditions, but it remains less clear whether operating-condition-induced domain differences can be reduced through transfer learning. This thesis addresses that gap by combining IC-informed feature engineering, explicit cross-condition domain definition, model selection using Optuna with the TPE sampler, and fine-tuning-based transfer learning to study both diagnostic SOH estimation and prognostic RUL prediction under heterogeneous constant-current operating conditions.

2

Theory

This chapter introduces the theoretical background that supports the modelling choices in this thesis. The focus is not on the full battery health literature as a separate review chapter, but on the concepts that are directly needed for the later method design and interpretation of results. In particular, this chapter introduces the ideas behind battery diagnostics and prognostics, feature extraction from controlled constant-current cycling data, transfer learning for cross-condition prediction, hyperparameter optimization, and sensitivity analysis under limited early-life information.

2.1 Battery Diagnostics and Prognostics

2.1.1 State of Health Estimation

State of health (SOH) is a central concept in lithium-ion battery diagnostics because it provides a compact description of how far a battery has degraded relative to its rated condition. In controlled cycling studies, SOH is commonly linked to the available capacity of a cell and is therefore used to summarize long term ageing in a physically meaningful and convenient form [3, 7, 8]. When repeated reference measurements are available, SOH can be tracked along the life trajectory of a cell and used to compare degradation patterns across different batteries or operating groups.

In this thesis, the current capacity used for SOH labelling is obtained from a reference performance test (RPT). An RPT is a controlled diagnostic measurement inserted at selected points along the ageing trajectory, and such diagnostic tests are commonly used to connect measurable capacity and voltage responses with underlying degradation modes [2]. During such a test, the cell is first fully charged and then fully discharged under a specified protocol, so that the measured discharge capacity represents the capacity that is currently available at that ageing stage. This RPT-based capacity measurement provides the direct basis for assigning a SOH value to the corresponding observation point.

A common capacity-based definition of SOH is

$$\mathrm{SOH}_k = \frac{C_{\mathrm{meas},k}}{C_{\mathrm{nom}}} \times 100\%, \tag{2.1}$$

where $C_{\mathrm{meas},k}$ denotes the discharge capacity measured during the RPT at ageing stage k, and C_{nom} denotes the nominal capacity of the cell. Under this definition,

SOH decreases as degradation accumulates. The value of SOH does not directly describe how much useful life remains, but it indicates the current health level of the battery and therefore serves as a natural diagnostic target. Figure 2.1 illustrates this pointwise SOH labelling process together with the RPT-based capacity measurement used to obtain each label.

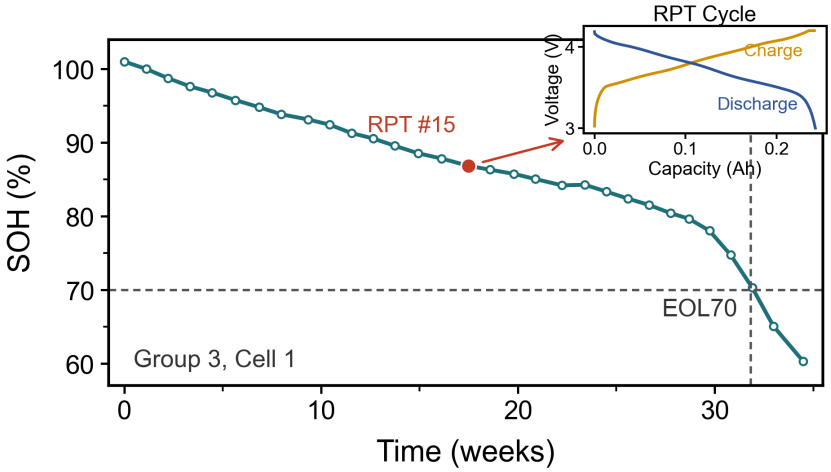


Figure 2.1: Conceptual comparison of state of health (SOH) and remaining useful life (RUL) definitions with an RPT-based observation inset. SOH labels can be assigned to multiple observations along the degradation trajectory from measured RPT capacity, whereas RUL is defined as the remaining time from a selected early observation week to the end-of-life (EOL) threshold defined at 70% SOH. The figure is used here to clarify the definition of a pointwise health indicator obtained from a controlled capacity measurement.

For data-driven modelling, SOH estimation is often formulated as a regression task in which measurable signals, engineered features, or latent representations are mapped to a continuous health value. In this thesis, SOH estimation also plays an additional role as an exploratory baseline. It is used to examine whether a relatively direct health indicator is sufficient to expose differences between source and target battery groups before moving to the more demanding task of remaining useful life (RUL) prediction.

2.1.2 Remaining Useful Life Prediction

RUL is a prognostic quantity rather than a diagnostic one. Instead of describing the current condition only, it estimates how much usable life is left before a predefined end-of-life (EOL) criterion is reached. In battery applications, this criterion is usually expressed through a degradation threshold, for example a chosen SOH level below which the cell is considered no longer suitable for the intended use.

A general RUL definition can be written as

$$\mathrm{RUL}_k = t_{\mathrm{EOL}} - t_k, \tag{2.2}$$

where t_k is the current time or cycle index and t_{EOL} is the time or cycle index at which the EOL threshold is reached. Compared with SOH estimation, RUL prediction

is usually more difficult because it depends not only on the present degradation state but also on the future degradation trajectory. This means that the model must capture patterns that are predictive of remaining lifetime rather than only reconstructive of current health.

In cross-condition settings, RUL is especially informative because different battery groups may show similar health values at a given observation point while still having different future ageing behaviour. For this reason, RUL prediction is often a stricter test of model generalization and domain adaptation than SOH estimation. Figure 2.2 illustrates how SOH samples and RUL samples are constructed under multiple operating conditions, emphasizing that RUL converts an observed battery state into a future-oriented lifetime target.

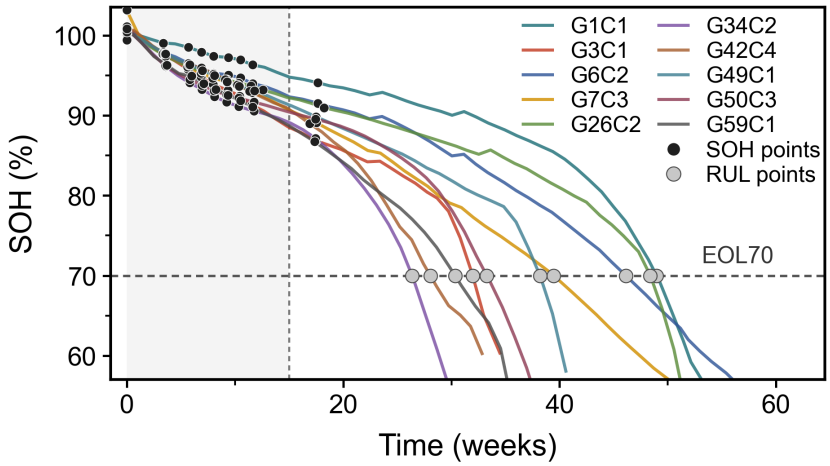


Figure 2.2: Illustration of state of health (SOH) and remaining useful life (RUL) sample mapping under multiple operating conditions. SOH labels are assigned to observation points along degradation trajectories, whereas RUL is defined from a selected observation week to the end-of-life (EOL) threshold defined at 70% SOH. The figure clarifies how the same degradation process can be organized either as pointwise health estimation or as future-oriented lifetime prediction.

2.1.3 Relationship Between State of Health and Remaining Useful Life

SOH and RUL are closely related but not interchangeable. SOH summarizes current degradation, whereas RUL reflects the distance from the current state to a future failure threshold. Figure 2.1 shows this distinction at the level of a single degradation trajectory: SOH is assigned to observation points from RPT-based capacity measurements, while RUL is measured from a selected observation point to the EOL threshold. Figure 2.2 extends the same distinction to multiple operating conditions and shows how the sample mapping changes when the target is changed from current health estimation to future lifetime prediction.

If degradation were perfectly linear and identical across all cells, SOH could almost directly determine RUL. In practice, however, degradation rates vary across cells, operating conditions, and ageing stages. Two cells with similar SOH can

therefore have different remaining life. This is why the two figures should be read together: Figure 2.1 explains how the diagnostic SOH label is obtained, whereas Figure 2.2 explains how that health trajectory is reformulated into RUL samples for prognostic modelling.

This distinction is important for the present thesis. SOH provides a direct description of the current health condition of a battery and is therefore a natural target for diagnostic analysis. RUL, in contrast, reformulates the same degradation process into a prediction problem tied to the remaining distance to EOL. In this sense, SOH and RUL do not compete with one another; rather, they emphasize different aspects of battery health modelling. The present thesis uses SOH as a preliminary diagnostic target and RUL as the final prognostic target in the later transfer learning analysis.

Figure 2.2 also helps explain why the final transfer learning task is shifted from SOH estimation to RUL prediction. In a SOH-based setting, labels are attached to observation points distributed along the degradation trajectory, so the resulting supervised samples still mainly describe the current state of the battery at each observation. By contrast, the RUL formulation maps an early observation to the remaining distance to the EOL threshold defined at 70% SOH, which makes the prediction target more directly connected to future degradation separation. For the present thesis, this is important because the study is not only concerned with pointwise health diagnosis, but with whether knowledge can be transferred across battery groups in a way that remains informative for prognosis. RUL is therefore adopted as the final transfer learning target because it provides a more suitable perspective for evaluating cross-condition predictive usefulness than SOH alone.

This theoretical distinction also creates a natural transition to the next part of the chapter. Once SOH and RUL have been separated conceptually, the key methodological question becomes how the underlying cycling data should be represented so that both tasks can be studied consistently. The following section therefore turns to IC-based feature representation, which provides the common input basis for the preliminary SOH analysis and the later week-based RUL transfer learning framework.

2.2 Feature Representation from Battery Cycling Data

In this thesis, the engineered inputs are defined as a ten feature general family that is used consistently throughout the later SOH and RUL analyses. This family can be organized into two broad categories: RPT-derived features and cycling condition features.

2.2.1 Reference Performance Test-Derived Features

Feature engineering is an important step in battery health modelling because raw cycling data are high-dimensional and often contain redundant or noisy information. A useful feature representation should preserve degradation-relevant patterns while

reducing complexity for the learning algorithm. In the present thesis, the first broad feature category is formed by features extracted from repeated reference performance tests (RPTs). As summarized in Table 2.1, this category includes the IC-based descriptors $f1$, $f7$, and $f9$, the DVA-based capacity change descriptors $f4$ and $f8$, and the CV hold time descriptors $f2$ and $f10$.

Incremental capacity is usually expressed as

$$\mathrm{IC}(V) = \frac{\mathrm{d}Q}{\mathrm{d}V}, \tag{2.3}$$

where Q denotes charge quantity or capacity and V denotes voltage. Peaks, shifts, and shape changes in IC curves can reflect changes in internal cell behaviour and therefore provide information about ageing. When measurements are collected under controlled cycling protocols, IC features can be especially useful because the input signals are more comparable across cycles and cells.

In practice, the RPT-derived category does not rely only on the visual interpretation of full IC curves. The same RPT framework can also be transformed into numerical descriptors by sampling characteristic IC regions, summarizing DVA-based capacity changes, and tracking CV hold time behaviour. Together, these variables describe how the measured voltage and capacity response evolves with ageing and therefore provide a common representation of battery state change under repeated cycling tests.

2.2.2 Cycling Condition Features

The second broad feature category contains cycling condition features. These variables describe the operating severity applied to the cell rather than the voltage response directly, and in this thesis they correspond to $f3$ (DoD), $f6$ (charging C-rate), and $f5$ (a combined charge induced stress indicator). Such quantities are important because degradation is influenced not only by the current measured response of the battery but also by the conditions under which the battery has been cycled.

Table 2.1: Ten feature general family used for subsequent feature engineering and selection. The features are organized into two broad categories: RPT-derived features and cycling condition features. Later state of health (SOH) and remaining useful life (RUL) experiments use subsets selected from this common feature family.

ID	Feature	Broad Category	Interpretation
f1	$\log(\mathrm{mean}(\Delta dQ/dV)_{w3-w0}^{3.6\mathrm{V}-3.9\mathrm{V}})$	RPT-derived feature	mean IC change over the 3.6–3.9 V region
f2	$\log(\lvert\Delta\mathrm{CV\ Time}_{w3-w0}\rvert)$	RPT-derived feature	change in CV hold time from week 0 to week 3
f3	DoD	Cycling condition feature	depth of discharge
f4	ΔQ_{w3-w0}^{1}	RPT-derived feature	DVA-based capacity change descriptor #1
f5	$C_{\mathrm{chg}}^{0.5}\mathrm{DoD}^{0.5}$	Cycling condition feature	combined charge induced stress indicator

Table 2.1: Ten feature general family used for subsequent feature engineering and selection (continued).

ID	Feature	Broad Category	Interpretation
f6	C_{chg}	Cycling condition feature	charging C-rate
f7	$\log(\mathrm{var}(\Delta dQ/dV)_{w3-w0}^{3.0\mathrm{V}-3.6\mathrm{V}})$	RPT-derived feature	variance of the low voltage IC feature
f8	ΔQ_{w3-w0}^{3}	RPT-derived feature	DVA-based capacity change descriptor #3
f9	$\log(\lvert\mathrm{mean}(\Delta dQ/dV)_{w3-w0}^{3.0\mathrm{V}-3.6\mathrm{V}}\rvert)$	RPT-derived feature	mean of the low voltage IC feature
f10	$\log(\lvert\Delta\mathrm{CV\ Time}_{w0}\rvert)$	RPT-derived feature	CV hold time descriptor of the initial RPT

Taken together, the RPT-derived features and the cycling condition features define the ten feature general family used throughout this thesis. All subsequent feature engineering and feature selection are performed within this family rather than by introducing unrelated feature types at later stages. This keeps the comparison between SOH and RUL tasks methodologically consistent while still allowing different subsets to be selected for different experiments. Table 2.1 summarizes the full feature family and serves as a reference for the discussions in Sections 2.2.1 and 2.2.2.

This organization makes the role of the feature family more explicit. Features in the first category primarily describe how the electrochemical response changes during controlled RPT measurements, whereas features in the second category summarize the applied operating severity. For SOH estimation, these features help characterize the current degradation state. For RUL prediction, they additionally provide information about how present response and operating history may relate to future degradation trajectory.

2.3 Transfer Learning for Battery Health Prediction under Varying Operating Conditions

2.3.1 Definitions of Domain, Task, and Transfer Learning

In the standard machine learning formulation, a domain D contains two components: a feature space $\mathcal{X}$ and a marginal probability distribution $P(X)$ over that space. A task T contains a label space $\mathcal{Y}$ together with a predictive function $f(\cdot)$, or equivalently the conditional relation that maps inputs to outputs. Under this formulation, the domain describes what kind of data are observed and how they are distributed, whereas the task describes what prediction problem is to be solved.

For battery health applications, the domain mainly refers to the measured features and their distribution, while the task refers to the battery health quantity to be estimated or predicted. For example, a domain may be built from IC-based, DVA-based, CV time, and cycling condition features collected from one cell or one battery group, whereas the task may be SOH estimation or RUL prediction. The

same task can therefore be studied under different domains, and the same domain can in principle support different tasks.

Given a source domain D_s and source task T_s, transfer learning aims to improve the predictive function in a target domain D_t and target task T_t by using knowledge learned from the source setting. In the most general case, transfer learning is relevant when the source and target differ in domain, task, or both. In the present thesis, the main interest is supervised battery health regression, where source and target tasks are closely related and the central question is whether knowledge learned from one battery setting can improve prediction in another related setting with limited target data.

2.3.2 Domain Shift in This Thesis

In this thesis, domain shift is defined in two ways. The first appears in the single cell SOH setting. A single battery cell provides many SOH labelled observations along its degradation trajectory, so one subset of observation points can be assigned to the source side and another disjoint subset can be assigned to the target side. In this case, the source and target are constructed from different parts of the available SOH samples, and the resulting shift is relatively mild because the data still come from the same cell trajectory.

The second definition appears in the cross-condition setting. Here, one set of battery cells is treated as the source domain and another disjoint set of battery cells is treated as the target domain. This creates a stronger domain difference because the source and target no longer differ only by sampled observations, but also by cell-to-cell variation and, in some protocols, by variation across operating conditions. From the perspective of this thesis, this second case is the more demanding and practically relevant one, because it better reflects the problem of transferring predictive knowledge from one battery group to another.

2.3.3 Fine-Tuning for Transfer Learning

A practical strategy for transfer learning in regression is fine-tuning. In this approach, a model is first trained on the source domain so that it learns a useful parameter initialization. The learned parameters are then adapted using a smaller amount of labelled data from the target domain. This differs from training a new model from the beginning, because the optimization starts from a representation that already contains task relevant structure.

Fine-tuning is attractive for battery prognostics for several reasons. First, it provides a simple way to combine source domain knowledge with limited target domain supervision. Second, it allows the model to retain shared degradation patterns while still adjusting to target specific behaviour. Third, it is compatible with common neural network training pipelines and can therefore be integrated with modern hyperparameter optimization frameworks.

Conceptually, fine-tuning assumes that the source and target tasks are related closely enough for parameter reuse to be meaningful. If the domains are too different, the transferred representation may not help and can even harm performance. The

later empirical sections of this thesis therefore evaluate whether fine-tuning improves cross-condition RUL prediction relative to non transfer baselines.

2.4 Model Selection and Hyperparameter Optimization

2.4.1 Hyperparameter Optimization

Hyperparameter optimization (HPO) refers to the process of selecting a hyperparameter configuration that minimizes the validation or generalization error of a learning algorithm. Formally, it can be viewed as an outer loop optimization problem over a hyperparameter search space, where each candidate configuration is evaluated by training the model and measuring its predictive loss on validation data [36]. Hyperparameters are quantities that control the structure or training behaviour of a learning algorithm but are not learned directly from the data during parameter estimation. Model depth, hidden dimensions, learning rate, batch size, regularization strength, and training duration are typical examples.

In this thesis, hyperparameter optimization is needed because model performance depends not only on the selected features and target definitions but also on how the learning algorithm is configured. Manual tuning can be inefficient and subjective, especially when several hyperparameters interact with one another. A configuration that performs well for one battery group or one task may be suboptimal for another. Because this thesis compares baseline and transfer learning settings under multiple cross-condition scenarios, a systematic search strategy is preferable to ad hoc trial and error.

2.4.2 Optuna with the TPE Sampler

Hyperparameter optimization (HPO) aims to identify the hyperparameter configuration that minimizes the validation loss. Given a hyperparameter configuration $\lambda \in \Lambda$, a training set D_{train}, a validation set D_{val}, and a learning algorithm A_λ, the HPO problem is formulated as

$$\lambda^* = \arg\min_{\lambda \in \Lambda} \mathcal{L}_{\text{val}} \left(A_\lambda(D_{\text{train}}), D_{\text{val}} \right), \tag{2.4}$$

where λ^* is the optimal hyperparameter configuration, Λ is the search space, and $\mathcal{L}_{\text{val}}$ denotes the validation loss.

In this work, Optuna is used as the optimization framework, and the Tree-structured Parzen Estimator (TPE) is adopted as the sampler. Let y denote the objective value obtained from a trial. Instead of directly modelling $p(y \mid \lambda)$, TPE models $p(\lambda \mid y)$ by separating previous trials using a threshold y^*:

$$p(\lambda \mid y) = \begin{cases} l(\lambda), & y < y^*, \\ g(\lambda), & y \geq y^*, \end{cases} \tag{2.5}$$

where $l(\lambda)$ is the density estimated from good trials and $g(\lambda)$ is the density estimated from the remaining trials. The threshold y^* is usually defined by a quantile of the observed objective values, with

$$\gamma = p(y < y^*). \tag{2.6}$$

For a minimization problem, the expected improvement is given by

$$\mathrm{EI}_{y^*}(\lambda) = \int_{-\infty}^{y^*} (y^* - y) p(y \mid \lambda)\, dy. \tag{2.7}$$

Using Bayes' rule and the TPE formulation, the expected improvement can be expressed as

$$\mathrm{EI}_{y^*}(\lambda) \propto \left(\gamma + (1-\gamma)\frac{g(\lambda)}{l(\lambda)}\right)^{-1}. \tag{2.8}$$

Thus, maximizing the expected improvement is equivalent to favouring configurations with a high ratio of $l(\lambda)$ to $g(\lambda)$:

$$\lambda_{\text{next}} = \arg\max_{\lambda \in \Lambda} \frac{l(\lambda)}{g(\lambda)}. \tag{2.9}$$

Therefore, TPE guides the search toward regions of the hyperparameter space that are more likely to produce low validation loss.

2.5 Sensitivity Analysis Under Limited Early-Life Data

2.5.1 Motivation for Sensitivity Analysis

In practical battery prognostics, prediction performance is strongly affected by how much information is available at the time of decision making. A model that performs well after observing a large portion of the degradation trajectory may be much less useful if only early-life data are available. For this reason, it is important to evaluate not only overall prediction accuracy but also sensitivity to data availability.

Sensitivity analysis in this context refers to a structured examination of how modelling performance changes when the observable data window is restricted. Rather than treating all input availability conditions as equivalent, this analysis makes explicit whether the proposed workflow remains informative when the model must predict long before the battery approaches EOL.

2.5.2 Early Stage Information and Prediction Difficulty

Early-stage prediction is intrinsically difficult because many degradation trajectories remain weakly separated in the initial part of life. At that stage, measured differences between cells may be subtle, noisy, or only indirectly related to long term failure timing. As a result, the same modelling framework can display different strengths depending on whether it is applied to early, middle, or late observations.

For a transfer learning study, this issue is particularly important. If source domain knowledge is genuinely useful, it may provide larger benefits when target domain information is scarce. Conversely, if the target domain has sufficient data, the relative value of transfer learning may decrease. Sensitivity analysis therefore helps interpret not only absolute prediction performance but also the conditions under which transfer learning is most beneficial.

2.5.3 Sensitivity to Fine-Tuning Battery Selection

In a fine-tuning setting, performance may depend not only on how many target domain batteries are available, but also on which specific batteries are chosen for adaptation. When only a small number of batteries are used for fine-tuning, the selected subset may not fully represent the variation of the target domain. Some batteries may reflect typical degradation behaviour, whereas others may emphasize less common trajectories or condition dependent effects. As a result, the apparent benefit of transfer learning can vary across different fine-tuning subsets even when the model architecture and training procedure remain unchanged.

For this reason, it is theoretically meaningful to examine sensitivity to fine-tuning battery selection. In this thesis, such sensitivity can be studied by using different random seeds to generate different subsets of fine-tuning batteries and then evaluating how strongly the final prediction performance changes across those selections. This analysis serves as a robustness check on the transfer learning workflow: if the method is stable, performance should remain reasonably consistent across different seed-based battery selections; if it is unstable, large variation would indicate that conclusions depend strongly on which few target batteries were used for adaptation. The analysis therefore provides theoretical support for interpreting fine-tuning results under limited target domain data.

2.6 Summary of Theoretical Background

The theoretical background of this thesis can be summarized briefly as follows. SOH estimation and RUL prediction describe two related but distinct battery health tasks, and the ten feature family introduced in this chapter provides a common input representation for both. Transfer learning is defined through domain and task relationships, and in this thesis the relevant domain shift appears both within single cell SOH observations and across different battery cells, with fine-tuning used as the main adaptation strategy. Hyperparameter optimization, implemented later with Optuna using the TPE sampler, provides a systematic way to configure the learning models. Finally, sensitivity analysis is used to interpret performance under limited early-life information and under different fine-tuning battery selections. Together, these concepts form the theoretical basis for the methods developed in the next chapter.

3
Methods

This chapter describes how the battery data are organized, how the prediction targets are defined, and how the modelling workflow is constructed for both the preliminary state of health (SOH) task and the following remaining useful life (RUL) transfer learning task. The overall strategy is to begin with a relatively direct diagnostic setting, then move to a more challenging prognostic setting in which cross-condition domain differences become more visible. To keep the comparison coherent, the same general feature family is used throughout the study, while the target definition, feature selection (selection happens in the general feature family), training protocol, and evaluation setting are adjusted according to the modelling objective.

3.1 Data and Preprocessing

3.1.1 Dataset Description

The data used in this thesis come from the ISU-ILCC NMC/Gr battery aging dataset introduced in *Predicting battery lifetime under varying usage conditions from early aging data* [18]. This dataset contains controlled cycling experiments on lithium-ion battery cells operated under varying usage conditions, and it is publicly available through Iowa State University DataShare [40]. The grouped structure of the dataset is especially important for this thesis because each group can be interpreted as a distinct but related domain under different charge C-rate, discharge C-rate, and depth of discharge settings.

At the experimental design level, the dataset was organized into 64 groups with four cells per group. In the released data, however, the full group G15C1–G15C4 and one additional cell, G61C1, are missing. As a result, 251 cells from 63 represented groups are available for analysis. This point is important for interpreting the data split: the dataset is not simply a flat collection of battery trajectories, but a structured set of group wise ageing experiments under heterogeneous operating conditions.

Table 3.1: Compact summary of the ISU dataset used in this thesis. End-of-life (EOL) threshold rows use the corresponding state of health (SOH) level.

Item	Summary
Available batteries	251 cells
Covered experimental groups	63 groups
Nominal group design	64 groups with 4 cells per group

Table 3.1: Compact summary of the ISU-ILCC NMC/Gr dataset as used in this thesis (continued).

Item	Summary
Missing data in released set	G15C1–G15C4 and G61C1
EOL threshold defined at 70% SOH: label availability	247 cells
EOL threshold defined at 70% SOH: lifetime range	1.48–68.09 weeks
Charging C-rate range	0.5C–3.0C
Discharging C-rate range	0.5C–2.475C
DoD range	3.5%–99.7%

Table 3.1 summarizes the main dataset characteristics used in this thesis. Besides the total number of cells and groups, the table highlights the availability of early-life week 5 features, the number of RPT-based SOH samples, the availability of end-of-life (EOL) labels defined at 80% and 70% SOH, and the range of operating conditions. These summary statistics show that the dataset is broad enough to support both diagnostic SOH estimation and prognostic RUL modelling, while still preserving clear group level heterogeneity.

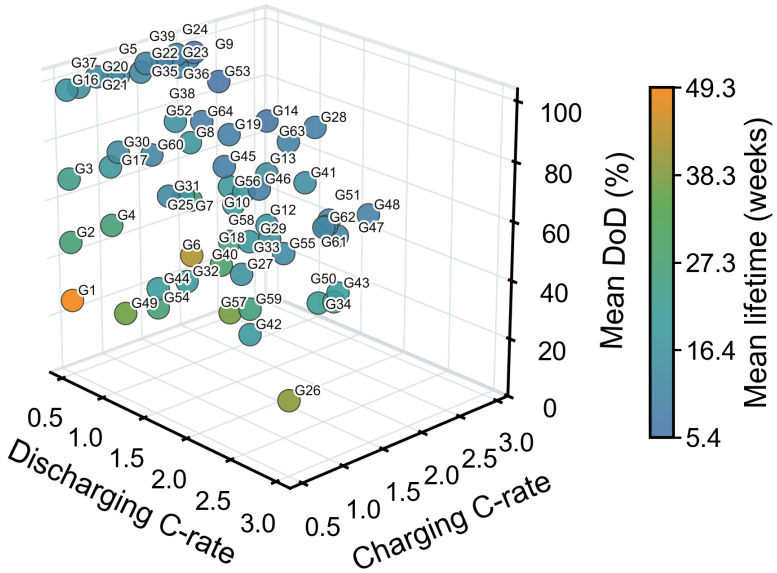


Figure 3.1: Operating condition distribution of the available battery groups. The cell groups are shown in the operating condition space defined by charging C-rate, discharging C-rate, and mean depth of discharge (DoD). Group level lifetime is used as the degradation outcome, illustrating that cells tested under different operating conditions can have substantially different degradation behaviour.

For each cell, the available measurements include repeated reference performance tests (RPTs) together with interpolated discharge information denoted as $Q_{\text{interpolated}}$. These measurements provide a consistent basis for constructing health-related features across the full degradation trajectory. In this thesis, the week index is used as the main time unit instead of equivalent full cycles (EFC) or raw cycle count. This

choice follows the structure of the released dataset and makes the later early-week RUL setting easier to define and interpret.

The dataset is therefore treated as a collection of cell wise trajectories organized by group rather than as isolated and independent points. Each trajectory contains a sequence of observations collected over the ageing process of one cell, while each group corresponds to a specific operating condition setting. Figure 3.1 visualizes this group level domain structure. This structure is essential for defining both the diagnostic target, which describes the current condition of the cell, and the prognostic target, which describes the remaining life before EOL threshold is reached.

3.1.2 Sample Construction

The difference between SOH samples and RUL samples should be clarified at the outset. As illustrated in Figure 2.2, SOH labels are attached to multiple observation points along one ageing trajectory. This means that a single battery can contribute many SOH samples taken at different stages of degradation. By contrast, in the week-based RUL setting used in this thesis, the label is defined from one selected early observation point to the EOL threshold defined at 70% SOH. Under this construction, each battery contributes one RUL sample for a given observation-week setting. Figure 2.2 therefore serves as the common visual reference for understanding why SOH is a pointwise health-estimation task with many points per battery, whereas RUL is a prediction task with one selected sample point per battery.

A modelling sample contains an input feature vector extracted from the corresponding RPT-related measurement and a target label determined according to the selected task.

For SOH estimation, the sample pair can be written as $(\mathbf{x}_k, y_k^{\mathrm{SOH}})$, where $\mathbf{x}_k$ is the feature vector obtained at observation k and y_k^{SOH} is the SOH label at the same observation. This formulation creates a direct mapping between the present measurement and the present health level.

For RUL prediction, the sample pair is written as $(\mathbf{x}_k, y_k^{\mathrm{RUL}})$, where the input represents the battery state at the selected early observation k, and the label corresponds to the remaining life until the cell reaches the predefined EOL threshold. Compared with SOH estimation, this task is more demanding because the model must infer future degradation progression from the currently battery behaviour.

Because multiple SOH samples can be generated from the same cell, whereas the week-based RUL setting contributes one selected sample per cell, careful split design is needed to avoid optimistic performance estimates. In this thesis, the data split is performed at the cell level rather than at the individual sample level. This means that all samples originating from a given cell are kept in the same subset, which reduces information leakage between training, validation, fine-tuning, and testing.

3.1.3 Feature Engineering

The input representation is based on a ten feature general family extracted from the interpolated discharge measurements, the corresponding incremental capacity and DVA responses, CV hold time behaviour, and key cycling condition variables.

The purpose of the feature engineering step is to convert the raw or semi processed battery measurements into a fixed length representation that is informative for both SOH estimation and RUL prediction.

Starting from the available $Q_{\text{interpolated}}$ measurements, the IC representation is derived from the voltage and capacity relationship according to

$$\mathrm{IC}(V) = \frac{\mathrm{d}Q}{\mathrm{d}V}. \tag{3.1}$$

In practice, the derivative is obtained numerically from the interpolated data. If necessary, smoothing or interpolation is applied before differentiation to reduce numerical noise and to ensure that different cells are compared on a common voltage basis. The resulting IC curves are then sampled or resampled over a shared voltage interval to obtain feature vectors of consistent dimension.

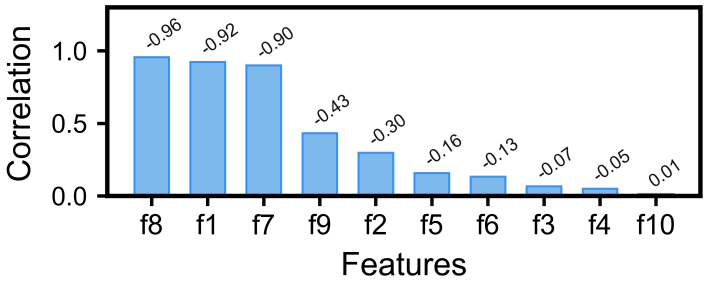


Figure 3.2: Overall state of health (SOH) correlations of the ten engineered features. Pearson correlations between each engineered feature and SOH are calculated across valid RPT samples. The reported SOH estimation experiments then deliberately use only feature $f1$ as a simple and interpretable diagnostic baseline.

The same ten feature family defines the common candidate pool across both experiments. This decision makes the comparison between tasks more meaningful, because changes in performance can then be interpreted mainly as a consequence of the target definition and the training protocol rather than as a consequence of switching to a fundamentally different input representation. In the reported SOH experiments, however, the ten feature family is first examined only as a candidate pool, and the actual SOH estimation input is restricted to the single feature $f1$. The RUL experiments later use selected subsets from the broader feature family. The guiding principle remains the same: the selected variables should retain degradation-related information while remaining numerically stable for model training.

Table 3.2: Single feature input definition used for the reported state of health (SOH) estimation experiments.

SOH analysis	Model	Feature	Feature definition	Feature count
Single cell estimation	Ridge	$f1$	Mean incremental capacity (IC) over 3.6–3.9 V	1
Cross-condition estimation	Ridge / Multilayer Perceptron	$f1$	Mean incremental capacity (IC) over 3.6–3.9 V	1

Figure 3.2 summarizes the SOH-oriented feature-correlation screening, while Table 3.2 specifies the exact input feature used in the reported SOH experiments. These two components should be interpreted together. Although the correlation analysis considers the full set of ten engineered features, the final SOH modelling setting is deliberately simplified: both the single-cell SOH estimation and the cross-condition SOH estimation use only $f1$, implemented as `feature_mean_ic`, as the model input. This design avoids unnecessary feature redundancy and focuses the SOH experiments on evaluating whether a representative IC-based feature can support health estimation across cells and operating conditions.

Figure 3.3 provides the corresponding feature-correlation analysis for the RUL prediction task. Compared with SOH estimation, RUL prediction depends more strongly on the selected feature representation because of the larger lifetime variation and cross-condition domain divergence. Therefore, Table 3.3 reports the final input-feature subset and the associated MLP architecture used in the main rerun400 transfer-learning experiment. Together, the RUL figure and table show how the candidate engineered features are reduced to the final model input configuration and combined with the selected neural-network architecture.

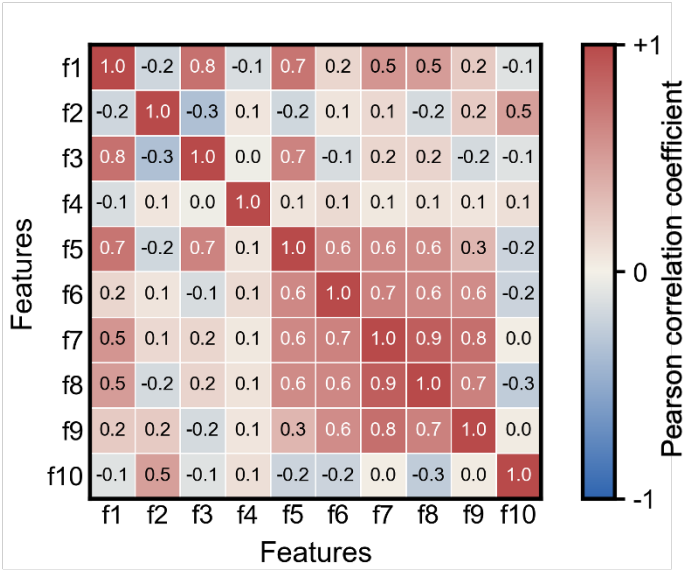


Figure 3.3: Correlation structure of the candidate input features for the main three stage remaining useful life (RUL) pipeline. A Pearson correlation matrix is used to visualize the relationships among the engineered early-life features for the RUL task with the end-of-life (EOL) threshold defined at 70% state of health (SOH). The matrix highlights redundant and complementary inputs before selecting the final feature subset for the reported main experiment.

The hidden-dimension setting reported in Table 3.3 refers to the internal architecture of the MLP used in the transfer-learning experiment. Specifically, the notation indicates one MLP model consisting of four hidden layers, with each hidden layer containing 64 neurons. Therefore, this setting should not be interpreted as four independent MLP models or as an ensemble of separate networks. Instead, it represents a single transfer model with a fixed four-layer hidden representation, where

the same architecture is used during source-domain pre-training and subsequent target-domain fine-tuning. This clarification is important because the reported performance corresponds to one unified neural network architecture, rather than the combined or averaged result of multiple models.

Table 3.3: Final input features and MLP architecture used in the main rerun400 remaining useful life (RUL) transfer learning experiment.

Item	Meaning / value	Scope	Role
$f1\_w5$	$f1$ feature extracted at week 5	week 5	main rerun400 input
$f6\_w5$	$f6$ feature extracted at week 5	week 5	main rerun400 input
Hidden dims	64, 64, 64, 64	MLP	transfer model architecture

3.1.4 Domain Definition for Modelling

After feature extraction, the samples are organized into modelling datasets according to cell group, task, and data role. Each cell group is treated as one domain. For a given transfer experiment, one group acts as the source domain and another group acts as the target domain. In this way, the domain definition is aligned with the practical objective of studying how well a model trained on one group can generalize to or be adapted for another group.

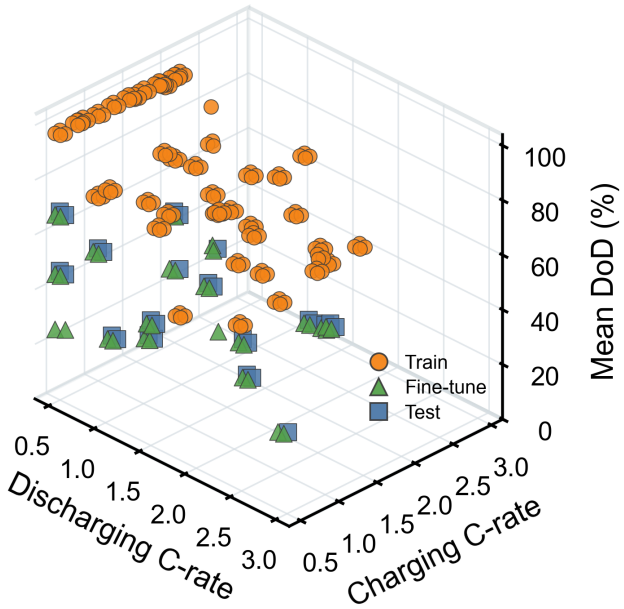


Figure 3.4: Cross-condition domain split used in the main observation week 5 remaining useful life (RUL) transfer learning experiment. Source domain training groups, target domain fine-tuning groups, and target-test groups are shown in the same operating condition space for the RUL task with the end-of-life (EOL) threshold defined at 70% state of health (SOH). The split is defined at the group/cell level so that target-test evaluation measures cross-condition generalization rather than random interpolation within the same operating condition.

Before model training, the feature vectors are standardized using statistics computed from the corresponding training data only. The same transformation is then applied to the associated validation, fine-tuning, and test subsets. This prevents the use of information from the evaluation data during preprocessing and ensures that all reported results reflect a fair predictive setting.

The target domain data are further divided according to their intended use. A small subset of target cells is used for fine-tuning or validation, while the remaining target cells are held-out for final testing. This distinction is especially important in the transfer learning experiments, where the aim is to evaluate whether limited target domain information is sufficient to adapt the source trained model to the new group.

The resulting split is illustrated in Figures 3.4 and 3.5. Figure 3.4 shows how the source domain training cells, target domain fine-tuning cells, and target-test cells are separated in the operating condition space. Figure 3.5 complements this view by comparing the corresponding lifetime distributions, which helps indicate whether the target-test cells differ from the source and fine-tuning cells not only in operating conditions but also in degradation range.

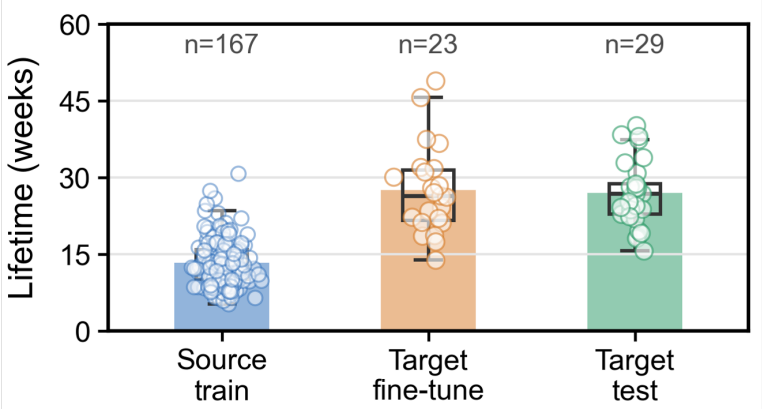


Figure 3.5: Lifetime distribution of source, target fine-tuning, and target-test cells. The distributions are compared using the end-of-life (EOL) threshold defined at 70% state of health (SOH). This diagnostic shows whether the selected partitions differ not only in operating conditions but also in lifetime range, which is important because a transfer learning task becomes more demanding when the target-test cells occupy a different lifetime region from the source or fine-tuning cells. The figure therefore supports the later interpretation of target-test remaining useful life (RUL) errors and fine-tuning performance.

3.2 Data-Driven Modelling

Two regression settings are used in this thesis so that each task can be evaluated with an appropriate level of model complexity. The preliminary SOH estimation experiments use Ridge regression as a transparent and stable baseline for matched domain and cross-condition diagnosis. The main cross-condition RUL experiments, by contrast, use a neural network regressor with tunable structure and training

hyperparameters so that fine-tuning and automated hyperparameter optimization can be applied in a consistent way.

This separation reflects the different roles of the experiments. The SOH tasks are used to establish whether the engineered features are diagnostically informative under simple modelling assumptions, whereas the RUL task is treated as the main transfer learning problem and therefore requires a more flexible model family. In all cases, the data split is defined at the cell level so that samples from the same cell are not distributed across training, validation, and test subsets.

3.2.1 Single Domain State of Health Estimation

The first experiment is a within-domain SOH estimation task implemented with Ridge regression. In this setting, the training data, validation data, and test data all come from the same cell group, and the prediction target is the current SOH value of each valid observation. The experiment uses only the single SOH feature $f1$ introduced in Table 3.2 and keeps the data split at the cell level so that samples from one cell do not appear in multiple subsets.

Table 3.4 summarizes the protocol used in this baseline experiment. The purpose is to establish a diagnostic reference under relatively favorable conditions, where the model does not need to generalize across different groups and where the effect of the engineered feature representation can be examined with a simple linear regressor.

Table 3.4: Protocol definition for the single domain state of health (SOH) estimation baseline. The setting is intentionally simple so that the within-domain diagnostic value of the selected single SOH feature can be evaluated before introducing cross-condition shift.

Component	Definition
Prediction target	SOH at the current observation point
Input features	Single feature $f1$ selected from the ten feature family (Table 3.2)
Model	Ridge regression
Domain relation	Training, validation, and test data drawn from the same cell group
Split principle	Cell wise split to avoid leakage between subsets
Purpose	Establish a matched domain diagnostic baseline before cross-condition experiments

This setting answers a basic question: can the chosen IC-based representation capture the current health state of cells when the training and evaluation domains are matched? If the answer is yes, the experiment provides evidence that the feature extraction and regression pipeline are reasonable even under a simple linear model. It also offers a reference point for later experiments, where the domain mismatch makes the prediction problem more challenging.

3.2.2 Cross-Condition State of Health Estimation

The second experiment keeps SOH as the prediction target and retains the same Ridge regression baseline, but introduces a domain shift between training and test-

ing. A model is trained using labelled samples from a source cell group and is then evaluated directly on a target cell group without any domain adaptation or fine-tuning. The purpose of this setting is to examine whether the SOH task alone is sufficient to reveal cross-condition divergence.

This experiment serves as a baseline without transfer learning. By keeping the SOH target, the single input feature $f1$, and the regression model aligned with Section 4.1, the comparison isolates the effect of domain mismatch more clearly. If the source trained model performs well on the target group, it suggests that the domain gap is limited for the SOH task. If the performance degrades substantially, it indicates that the source representation is not directly transferable. In the present thesis, this experiment is especially important because it motivates the shift from SOH estimation to RUL prediction. The cross-condition SOH task is informative as a diagnostic benchmark, but it is not necessarily sensitive enough to capture later stage divergence between battery groups.

3.2.3 Cross-Condition Remaining Useful Life Prediction with Fine-Tuning-Based Transfer Learning

The main experiment of the thesis is cross-condition RUL prediction with fine-tuning. As shown in Figure 3.6, the experiment compares two modelling branches under the same source, target fine-tuning, and target-test split. The benchmark branch trains a multilayer perceptron (MLP) directly from the combined source-domain data and limited target fine-tuning data, and then evaluates the resulting model on the held-out target-test cells. This branch provides the non-transfer reference result.

The fine-tuning transfer branch is implemented in two stages. Let $\mathcal{D}_s$ denote the source domain and $\mathcal{D}_t$ denote the target domain. The staged workflow can be summarized as follows:

1. pretrain the model on $\mathcal{D}_s$ to learn a source-domain representation for RUL prediction, with Optuna using the TPE sampler to select the source-pretraining hyperparameters;

2. transfer the pretrained model to $\mathcal{D}_t$, fine-tune it using a limited target-domain subset, and evaluate it on the remaining target-test cells, with Optuna using the TPE sampler again to select the fine-tuning hyperparameters.

This procedure is designed to test whether transferable degradation information exists across operating conditions and whether limited target domain supervision can improve prediction accuracy relative to the benchmark without transfer learning [28, 29, 31, 34].

The fine-tuning setting also makes it possible to study the practical role of target domain data availability. Because only a limited number of target cells are used for adaptation, the model must balance two competing goals: retaining useful source domain knowledge and learning the target specific characteristics that are not fully represented in the source group. The matched data split in Figure 3.6 is therefore important: both branches are evaluated on the same target-test data and with the

same metrics, so the final comparison isolates the effect of source pretraining and target fine-tuning rather than differences in the evaluation set.

Figure 3.6 should therefore be read as the experimental comparison schematic for the RUL task: the upper branch defines the non-transfer benchmark, the lower branch defines the source-pretraining and target fine-tuning workflow guided by Optuna with the TPE sampler, and the final comparison uses the same evaluation metrics, including MAE, MAPE, sMAPE, and wMAPE.

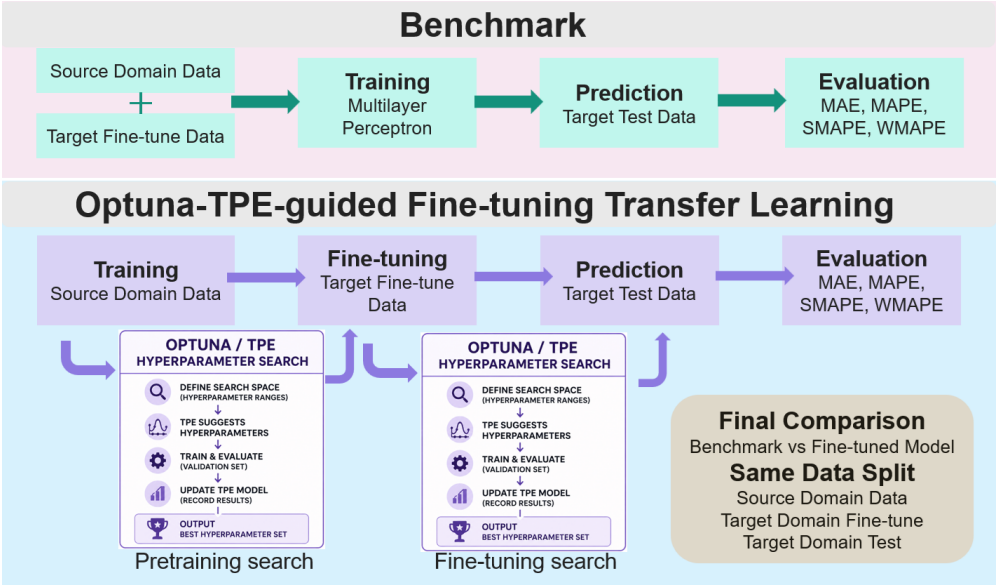


Figure 3.6: Integrated workflow for the cross-condition remaining useful life (RUL) prediction experiment. The benchmark branch trains a multilayer perceptron directly from the combined source-domain data and target fine-tuning data, and then evaluates the prediction on the target test data. The transfer-learning branch first trains the model on the source domain, then fine-tunes it with limited target-domain data, with Optuna using the TPE sampler for the staged hyperparameter searches in both the pretraining and fine-tuning stages. Both branches use the same data split and are evaluated with the same metrics, enabling a fair comparison between the benchmark model and the fine-tuned transfer model.

3.2.4 Hyperparameter Optimization

To reduce manual tuning and improve the robustness of model selection, hyperparameter optimization is performed with Optuna using the Tree-structured Parzen Estimator (TPE) sampler. The optimization objective is defined on a validation subset so that candidate configurations are selected according to predictive performance rather than subjective preference.

The search space includes both structural parameters and learning hyperparameters. Typical examples include the number of hidden layers, hidden dimensions, learning rate, batch size, regularization strength, dropout level, and the number of training epochs. Rather than fixing these settings by hand, the optimization procedure explores multiple candidate combinations and retains the configuration that achieves the best validation performance under the chosen metric.

The same general optimization logic is used for the baseline and transfer learning experiments, but the validation data are defined according to the task. In the transfer learning setting, the optimization is aligned with the target domain adaptation objective, so that the selected hyperparameters are relevant to cross-condition prediction rather than only to source domain fitting. This is particularly important because a configuration that is effective for source domain training alone may not be the best choice for subsequent fine-tuning on the target group.

The staged Optuna search with the TPE sampler is included in Figure 3.6. In this workflow, the search first identifies source-pretraining configurations, then optimizes target fine-tuning settings, and finally retrains and evaluates the selected configuration on the held-out target-test partition. This staged design keeps model selection aligned with the transfer learning objective rather than optimizing only for source domain fitting.

3.2.5 Evaluation Metrics

All prediction tasks in this thesis are treated as regression problems. To keep the evaluation consistent across the SOH and RUL experiments, the reported metrics are restricted to absolute error and percentage error measures: MAE, MAPE, sMAPE, and wMAPE [41]. These metrics are used because they describe prediction error directly, while avoiding additional goodness of fit or percentile-based summaries that are less central to the comparison in this thesis.

The mean absolute error (MAE) is used as the primary metric because it provides an interpretable measure of the average prediction deviation in the original target unit:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|. \tag{3.2}$$

The mean absolute percentage error (MAPE) expresses the absolute error relative to the true target value:

$$\mathrm{MAPE} = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right|. \tag{3.3}$$

Because percentage errors can behave differently when target magnitudes vary, two MAPE related variants are also reported. The symmetric mean absolute percentage error (sMAPE) normalizes the absolute error by the average magnitude of the true and predicted values:

$$\mathrm{sMAPE} = \frac{100}{n}\sum_{i=1}^{n}\frac{|y_i - \hat{y}_i|}{\left(|y_i| + |\hat{y}_i|\right)/2}. \tag{3.4}$$

The weighted mean absolute percentage error (wMAPE) aggregates the absolute errors before normalizing by the total magnitude of the true target values:

$$\mathrm{wMAPE} = 100\frac{\sum_{i=1}^{n}|y_i - \hat{y}_i|}{\sum_{i=1}^{n}|y_i|}. \tag{3.5}$$

In the tables, percentage-based metrics are reported in percent. Positive improvement values indicate that the adapted or fine-tuned model reduces the corresponding error relative to the benchmark model.

Beyond overall accuracy, the thesis also evaluates sensitivity to data availability. In the sensitivity analysis, the target domain information available for adaptation is restricted to the early part of the battery life trajectory. Several early-life observation windows are considered, and the transfer learning workflow is repeated under each window. By comparing the resulting performance levels, the study examines how strongly the effectiveness of cross-condition RUL prediction depends on the amount of early-stage information available for fine-tuning. This protocol complements the standard regression metrics by revealing not only whether the model works, but also under what information conditions it remains useful.

4

Results

4.1 Performance of Single Domain State of Health Estimation

The single domain state of health (SOH) estimation experiment is used as the first and most basic modelling step in the study. Before moving to cross-cell and cross-condition prediction, the initial aim is to examine the SOH prediction behaviour within one battery cell. This setting allows the relation between the selected IC-based feature $f1$ and the current SOH to be explored under a controlled matched-domain condition, where the training and test samples come from the same cell trajectory.

From a methodological perspective, this single-cell experiment has two roles. First, it verifies that the basic modelling pipeline is functional in a matched-domain setting. Second, it provides an intuitive starting point for later transfer learning analysis by beginning with the simplest case, where domain mismatch is intentionally minimized. The within-cell SOH task therefore acts as a reference case from which later domain shift experiments can be interpreted more clearly.

At the same time, the experiment also serves a more specific conceptual purpose: it provides a first check of how well SOH can be predicted when the data come from one continuous cell trajectory. The result suggests that the within-cell variation is not large enough to substantially reduce prediction accuracy in this representative case. This supports the use of the single-cell experiment as an initial baseline rather than as the main test of transfer-related difficulty.

Table 4.1: Representative single-run within-cell state of health (SOH) result for battery G1C1. The table reports the test-set metrics corresponding to the example shown in Figure 4.1.

Cell	n test	MAE	MAPE(%)
G1C1	11	0.0162	1.7531

For illustration, cell G1C1 is used as a representative example. Figure 4.1 shows the predicted-versus-true SOH relation for this within-cell experiment. The points remain close to the identity line, and the reported metrics (MAE = 0.016 and MAPE = 1.753%) confirm that the model can fit this matched-domain problem well. Table 4.1 reports the corresponding numerical result. Together, the figure and table show that the selected feature and regression pipeline can capture the

basic SOH trend within a single battery, which makes this experiment a useful introductory check before moving to stronger domain shift experiments.

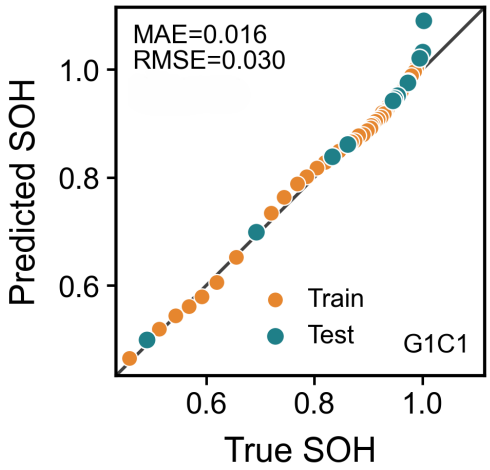


Figure 4.1: Predicted versus true state of health (SOH) for the within-cell prediction task on battery G1C1. A representative within-cell SOH experiment is shown using the single SOH input feature $f1$ and Ridge regression. Both training and test predictions lie close to the identity line, and the reported MAE = 0.016 and MAPE = 1.753% indicate that this matched-domain setting is easy to learn before any cross-condition transfer is introduced.

4.2 Cross-Condition SOH Estimation Performance

The cross-condition SOH estimation experiment is the first setting in which the train and test data are separated at the cell level. Unlike the single-cell setting in Section 4.1, the model is no longer trained and evaluated within one battery cell. Instead, it is trained on one subset of cells and evaluated on different held-out cells. This setting therefore provides a practical cross-cell benchmark for examining whether SOH prediction remains reliable when the test batteries and their degradation trajectories are not observed during training. In the following analysis, the results are interpreted from two complementary levels: the sample level, where all test observations are pooled together, and the cell level, where prediction errors are first calculated for each test cell and then summarized across cells.

Table 4.2: Overall target-test metrics for the cross-condition state of health (SOH) benchmark without transfer learning. The benchmark contains 684 samples from 45 cells distributed across 12 groups.

Model	n samples	n cells	n groups	MAE	MAPE (%)
Ridge	684	45	12	0.0486	5.8114
MLP	684	45	12	0.0626	8.1862

This experiment first provides a sample-level evaluation of the cross-cell SOH benchmark. Table 4.2 reports the overall target-test metrics after pooling all 684 SOH samples from 45 held-out cells. From this sample-weighted perspective, Ridge

regression performs better than the MLP, with MAE = 0.0486 and MAPE = 5.8114%, whereas the MLP gives MAE = 0.0626 and MAPE = 8.1862%. This indicates that the simpler linear baseline is sufficient to capture most of the SOH-related variation in this setting, even though the test cells are different from the training cells.

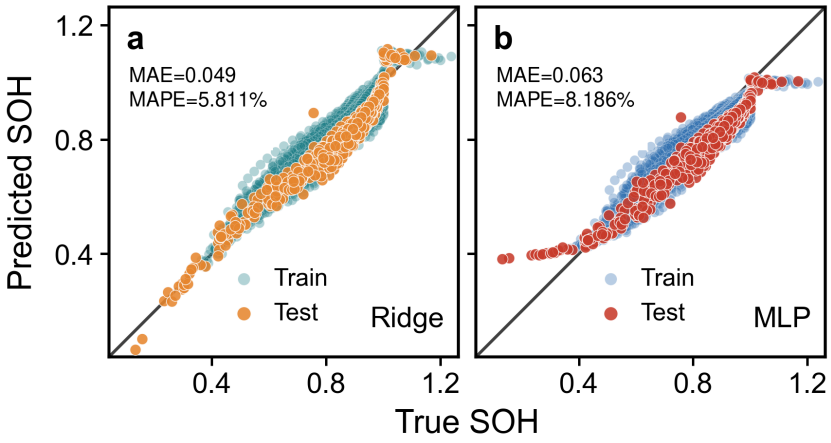


Figure 4.2: Predicted versus true state of health (SOH) for the cross-condition benchmark without transfer learning. Panel (a) shows Ridge regression and panel (b) shows the MLP under the same cross-condition training and test split. Ridge achieves MAE = 0.0486 and MAPE = 5.8114%, whereas the MLP gives MAE = 0.0626 and MAPE = 8.1862%. The tighter concentration of the Ridge predictions around the identity line indicates that the simpler model generalizes more robustly than the nonlinear alternative under cross-condition SOH domain shift.

Figure 4.2 provides the corresponding visual comparison between predicted and true SOH under the cross-condition SOH estimation setting. The Ridge regression predictions remain more concentrated around the identity line, indicating that the model can still capture the main SOH trend across different cells and operating conditions. In contrast, the MLP predictions show a wider spread and larger deviations from the identity line, suggesting weaker robustness under the same cross-condition split. This observation is consistent with the quantitative results, where Ridge regression performs more stably than the MLP baseline. However, the overall prediction pattern does not indicate a severe collapse of cross-cell SOH estimation. Although the cross-condition setting is more realistic and more challenging than the single-cell setting, most predictions still follow the general direction of the true SOH values. The errors increase compared with the matched or within-cell case, but the model does not completely fail when transferred across operating conditions. Therefore, the sample-level results suggest that the SOH task reveals some cross-condition difficulty, but it does not expose a strong and clearly separable domain divergence in the present dataset. This supports the interpretation that SOH estimation is useful as a preliminary diagnostic validation task, while the later RUL prediction task provides a more demanding setting for evaluating transfer learning.

Figure 4.3 then examines the same cross-cell experiment from the cell-level perspective. Here, the prediction error is first computed separately for each held-out

test cell, and the resulting cell-level errors are summarized by the mean and standard deviation. This analysis is important because the overall sample-level metrics in Table 4.2 are sample-weighted, meaning that cells with more SOH observations can have a larger influence on the final overall error. By contrast, the cell-level summary checks whether the conclusion remains consistent when the performance on individual batteries is considered more directly.

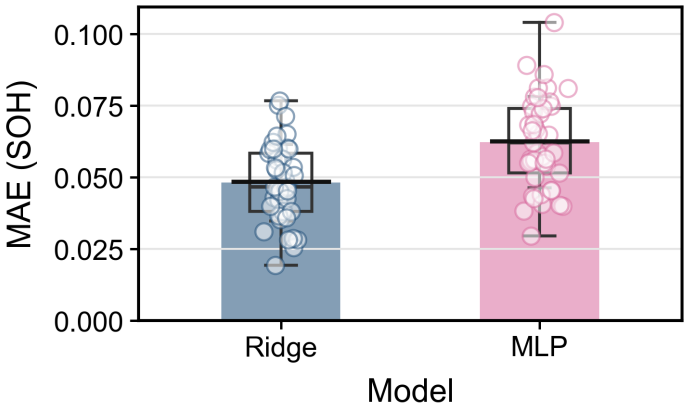


Figure 4.3: Cell level MAE summary for the cross-condition state of health (SOH) benchmark. The bars denote the cell level MAE mean, the error bars denote the cell level standard deviation, and the black horizontal lines denote the sample level overall MAE. Ridge shows both a lower average cell level error and a smaller error spread than the MLP, which confirms that its advantage is not limited to a few isolated cells but remains visible across the broader cross-condition evaluation.

Table 4.3 shows that Ridge again gives lower error than the MLP, with a cell-level MAE mean of 0.0482 and MAPE mean of 5.7473%. The MLP gives a higher cell-level MAE mean of 0.0623 and MAPE mean of 8.1406%, together with a larger MAPE standard deviation. The closeness between the black horizontal sample-level MAE lines and the cell-level MAE bars in Figure 4.3 further suggests that the overall result is not dominated by only a few test cells.

Table 4.3: Cell level error statistics for the cross-condition state of health (SOH) benchmark without transfer learning. Here, μ and σ denote the mean and standard deviation across cells.

Model	MAE min.	MAE max.	MAE mean	MAE std.	MAPE mean (%)	MAPE std. (%)
Ridge	0.0193	0.0767	0.0482	0.0133	5.7473	1.7751
MLP	0.0297	0.1042	0.0623	0.0159	8.1406	4.5945

Taken together, the cross-condition SOH results show that Ridge regression provides a more accurate and stable baseline than the MLP at both the sample level and the cell level. More importantly, the two-level analysis leads to the same interpretation: although the cross-cell setup is stricter than the within-cell experiment, the SOH task still does not reveal a strong enough domain divergence to clearly

demonstrate the value of transfer learning. The model errors vary across cells, but the variation remains relatively limited and the overall prediction trend is still reasonable. For this reason, Section 4.2 serves as a diagnostic transition rather than the final transfer learning task, motivating the later shift toward remaining useful life (RUL) prediction, where the target is more directly related to future degradation trajectory and end-of-life timing.

4.3 RUL Prediction with Transfer Learning Selected by Optuna with the TPE Sampler

The main RUL result is presented as a direct comparison between the benchmark without transfer learning and the transfer learning model selected by Optuna with the TPE sampler. Both models are evaluated on the same target-test partition under the end-of-life (EOL) threshold defined at 70% SOH, so the comparison focuses on whether the fine-tuning-based transfer workflow improves target condition RUL prediction rather than on the benchmark result in isolation.

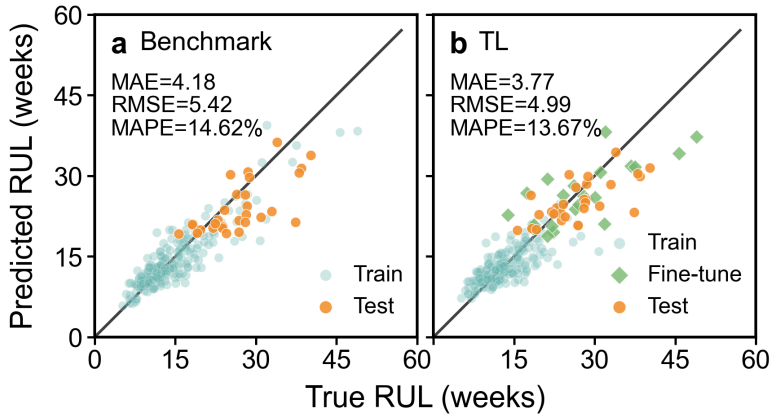


Figure 4.4: Predicted versus true remaining useful life (RUL) comparison between the benchmark and the transfer learning model selected by Optuna with the TPE sampler. Both models are evaluated on the same target-test cells using the RUL definition with the end-of-life (EOL) threshold defined at 70% state of health (SOH). Points closer to the identity line indicate better prediction calibration.

Figure 4.4 compares predicted and true RUL values for the benchmark and the transfer learning model selected by Optuna with the TPE sampler. The closer the points are to the identity line, the better the calibration of the prediction. The transfer learning result shows a tighter correspondence between predicted and true RUL values, indicating that source pretraining followed by target fine-tuning improves the model's ability to estimate remaining life on unseen target-test cells.

Table 4.4: Target-test metric comparison between the benchmark and the transfer learning model selected by Optuna with the TPE sampler. Positive improvement values indicate error reduction relative to the benchmark.

Metric	Benchmark	Optuna-selected transfer learning	Improvement
MAE	4.18	3.77	+9.8%
MAPE (%)	14.62	13.67	+6.5%
sMAPE (%)	15.91	14.11	+11.3%
wMAPE (%)	15.52	14.00	+9.8%

Figure 4.5 complements the parity plot by comparing the corresponding error distributions. This view shows not only the average performance difference, but also whether the transfer learning model reduces the spread of target-test errors. Together with the aggregate metrics in Table 4.4, the two figures show that the transfer learning workflow selected by Optuna with the TPE sampler consistently improves the main target-test RUL result relative to the benchmark.

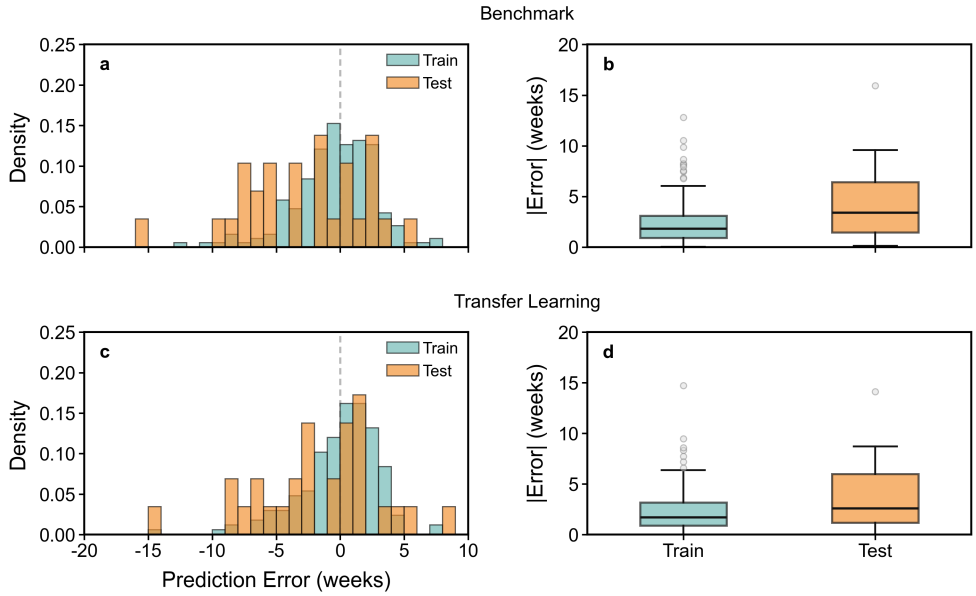


Figure 4.5: Target-test error distribution comparison between the benchmark and the transfer learning model selected by Optuna with the TPE sampler. The figure compares the spread and magnitude of remaining useful life (RUL) prediction errors under the same target-test protocol, showing whether fine-tuning reduces error dispersion as well as aggregate error.

4.4 Sensitivity and Robustness Analyses

The sensitivity checks were used to test whether the main RUL conclusion depends on the observation week and on the random selection of target domain fine-tuning cells. In both tables, the benchmark and fine-tuned models are evaluated on the corresponding target-test partition under the same RUL protocol with the EOL threshold defined at 70% SOH. Positive Delta values indicate that fine-tuning reduces the error relative to the benchmark, whereas negative values indicate worse performance after fine-tuning.

Table 4.5 summarizes the early-week sensitivity check. The week 5 result matches the main experiment reported in Section 4.3. Across weeks 5–10, fine-tuning improves MAE in four out of six settings, with the largest gain at week 10. However,

weeks 7 and 8 show negative transfer, indicating that later observation weeks do not automatically guarantee a better transfer effect. The week-based result therefore suggests that transfer performance depends not only on the amount of early-life information available, but also on how the target domain information aligns with the held-out target-test cells.

Table 4.5: Sensitivity of remaining useful life (RUL) prediction to the early observation week. Positive Delta values indicate error reduction from fine-tuning relative to the benchmark.

Week	Benchmark MAE	Fine-tuned MAE	ΔMAE (%)	Benchmark MAPE (%)	Fine-tuned MAPE (%)	ΔMAPE (%)
w5	4.18	3.77	+9.8	14.62	13.67	+6.5
w6	3.77	3.74	+0.8	13.80	13.35	+3.3
w7	4.53	4.57	-1.0	14.81	15.76	-6.4
w8	4.07	4.54	-11.7	13.49	16.93	-25.5
w9	5.06	4.70	+7.3	16.86	15.35	+9.0
w10	4.59	3.77	+17.9	15.13	11.77	+22.2

Table 4.6 reports the sensitivity to random target domain splitting. The first column gives a selection index from 1 to 21, where each index represents one target-domain fine-tuning/test split. Changing the selection changes which target cells are available for adaptation and which cells are used for final evaluation. The results show substantial variation across selections: some splits produce clear positive transfer, whereas others lead to negative transfer. This supports the interpretation that the benefit of fine-tuning is sensitive to the representativeness of the selected target fine-tuning cells.

Table 4.6: Sensitivity of remaining useful life (RUL) prediction to target domain fine-tuning/test selection. The selection index identifies the target-domain split used for fine-tuning and target testing. Positive Delta values indicate error reduction from fine-tuning relative to the benchmark.

Selection	Benchmark MAE	Fine-tuned MAE	ΔMAE (%)	Benchmark MAPE (%)	Fine-tuned MAPE (%)	ΔMAPE (%)
1	4.18	3.77	+9.8	14.62	13.67	+6.5
2	5.16	4.38	+15.2	18.15	15.81	+12.9
3	4.62	6.14	-32.9	17.17	23.52	-36.9
4	5.46	5.69	-4.2	17.52	17.80	-1.6
5	4.27	4.00	+6.5	14.62	13.63	+6.8
6	5.24	6.39	-22.0	20.44	26.61	-30.2
7	4.27	4.09	+4.4	15.98	16.72	-4.7
8	4.55	4.55	-0.1	17.62	18.77	-6.5
9	4.15	4.61	-10.9	15.69	16.31	-4.0
10	4.24	3.85	+9.3	17.29	15.19	+12.1
11	5.09	4.27	+16.2	19.53	16.76	+14.2
12	5.09	5.12	-0.6	19.29	20.07	-4.1
13	4.67	4.46	+4.6	18.57	15.92	+14.3
14	4.87	4.82	+1.1	16.92	16.50	+2.4
15	4.16	3.93	+5.5	15.35	14.52	+5.4
16	4.64	5.03	-8.5	16.64	17.35	-4.2
17	5.11	5.18	-1.3	17.88	19.21	-7.4

Table 4.6: Sensitivity of remaining useful life (RUL) prediction to target domain fine-tuning/test selection (continued).

Selection	Benchmark MAE	Fine-tuned MAE	ΔMAE (%)	Benchmark MAPE (%)	Fine-tuned MAPE (%)	ΔMAPE (%)
18	5.36	4.78	+10.8	21.44	18.97	+11.5
19	3.64	3.21	+11.8	13.72	12.60	+8.1
20	5.99	6.32	-5.5	21.62	22.86	-5.8
21	4.79	4.71	+1.6	15.70	15.92	-1.4

For the target-cell selection sensitivity check, the ΔMAE summary is $n = 21$, mean +0.5%, standard deviation 11.9%, median +1.6%, minimum −32.9%, and maximum +16.2%. Positive transfer is observed in 12 out of 21 selections. Thus, although the main configuration shown as selection 1 gives a clear improvement from fine-tuning, the broader selection sweep demonstrates that the transfer effect is not uniform across target domain splits.

5

Discussion and Conclusion

This chapter interprets the main findings of the thesis and discusses the factors that affect the reported transfer learning results. The focus is not only on the final prediction accuracy, but also on why the observed performance changes across state of health (SOH) estimation, cross-condition remaining useful life (RUL) prediction, feature week, domain definition, and target domain fine-tuning selection.

5.1 Summary of Main Findings

The SOH estimation experiments show that the selected IC-based representation is informative for battery health estimation, particularly in the matched-domain setting. In the representative within-cell experiment, the model predicts SOH accurately, confirming that the feature extraction and regression pipeline is functional. Under cross-condition SOH estimation, the task becomes more challenging, but Ridge regression remains more robust than the MLP baseline. This suggests that SOH estimation can expose domain shift, although it is still less demanding than the later RUL transfer learning task.

The main RUL experiment provides the central result of this thesis. With the EOL threshold defined at 70% SOH, the fine-tuned transfer learning model reduces the target-test error compared with the benchmark model. As summarized in Table 4.4, MAE decreases from 4.18 weeks to 3.77 weeks, corresponding to a 9.8% reduction. MAPE, sMAPE, and wMAPE also decrease. Together with the parity and error distribution comparisons in Figures 4.4 and 4.5, these results indicate that source-domain pretraining followed by target-domain fine-tuning can improve cross-condition RUL prediction.

However, the sensitivity checks show that the transfer gain is not fully stable. Table 4.5 shows that the benefit varies with the selected feature week, while Table 4.6 shows that it also depends on the split between target fine-tuning and target-test cells. Therefore, the thesis does not claim that transfer learning is uniformly beneficial in all configurations. Instead, the results suggest that fine-tuning can improve RUL prediction, but the magnitude and even the direction of improvement depend on the available target-domain information and the specific cross-condition split.

5.2 Interpretation of Feature Choice

The feature design used in this thesis is based on the idea that early-life measurements contain useful information about later degradation behaviour. Week-based

features are therefore practically attractive because they support RUL prediction before a large fraction of the cell lifetime has been observed. This is important for battery prognostics, where waiting for long degradation histories may limit the usefulness of a prediction model in real applications.

However, the sensitivity results also show that the selected feature week can influence both benchmark performance and transfer learning performance. A later feature week may provide richer degradation information, but it may also change the set of valid cells and the distribution of samples available for training, fine-tuning, and testing. As a result, feature week sensitivity cannot be interpreted only as a simple information gain effect. It also reflects changes in sample composition and in the relationship between the source, fine-tuning, and test subsets.

The final feature subset used in the main RUL experiment should therefore be interpreted as one reasonable and controlled configuration rather than as a universally optimal feature set. Different combinations of IC-based features, degradation summaries, or week-specific descriptors could change the stability of transfer. In particular, features that better capture early trajectory shape or uncertainty may help reduce the dependence on one specific observation week. This remains an important direction for future model development.

5.3 Operating Conditions and Material Effects

The operating condition space provides a useful first way to define battery domains. Charging C-rate, discharging C-rate, and depth of discharge directly affect cell ageing and therefore provide a physically meaningful basis for source and target separation. The cross-condition split used in this thesis follows this logic: cells are grouped according to experimental condition, and transfer learning is evaluated by testing whether information from one group can support prediction in another group.

Figure 5.1 extends the group level operating condition discussion by showing the actual cell level distribution used in the main experiment. Each point corresponds to one cell, with a small jitter added so that cells tested under the same nominal condition do not completely overlap. This view is important because the target domain fine-tuning cells and the held-out target-test cells are not arbitrary points in the full operating condition space. They occupy specific regions, which means that the value of fine-tuning depends partly on whether the selected fine-tuning cells provide useful support for the operating region represented by the target-test cells.

Nevertheless, similarity in operating conditions does not necessarily imply similar degradation behaviour. Two groups may appear close in the selected condition variables but still differ in lifetime distribution, degradation pathway, cell-to-cell variability, or material response. Conversely, groups that are separated in the operating condition plot may share degradation structure that is useful for prediction. This helps explain why cross-condition prediction is difficult and why the fine-tuning result can vary across target domain selections.

Material and dataset effects add another layer of complexity. The thesis uses the ISU-ILCC NMC/Gr dataset, and the learned relationships are shaped by the chemistry, test protocol, feature extraction procedure, and EOL definition used in

that dataset. Therefore, the reported transfer learning behaviour should be interpreted within this experimental context. Broader validation across additional chemistries, cell formats, and independent datasets would be needed before claiming general transfer performance across battery systems, because battery health literature repeatedly shows that ageing mechanisms, operating protocols, and feature observability can strongly affect SOH and RUL model behaviour [8, 14, 17].

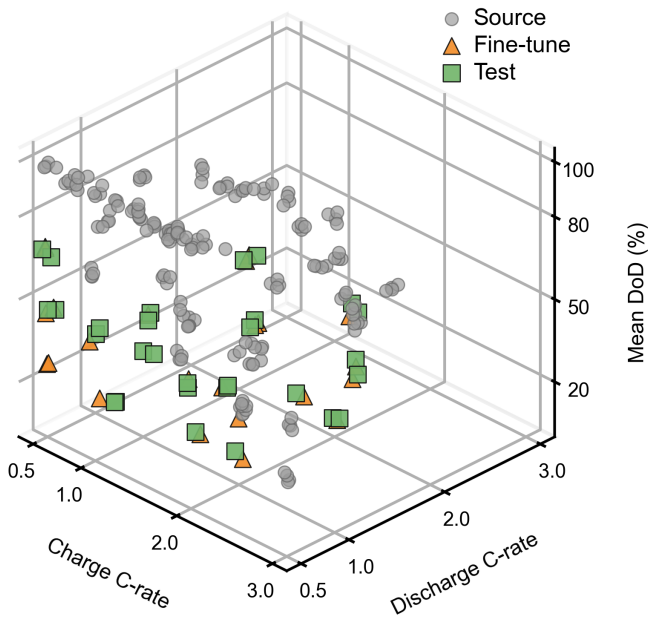


Figure 5.1: Cell level operating condition split used in the main remaining useful life (RUL) experiment. The source domain, target domain fine-tuning, and target-test cells are shown in the space defined by charge C-rate, discharge C-rate, and mean depth of discharge. A small jitter is applied only for visualization so that cells at the same nominal operating condition remain visible.

5.4 Limitations of Week Sensitivity Analysis

The week-based sensitivity analysis is useful because it shows that transfer performance changes when the early observation window changes. However, this analysis should not be interpreted as a pure test of which week is intrinsically best. When the feature week changes, the available cells and samples can also change because some cells may not have valid observations at later weeks or may reach the EOL threshold earlier. Therefore, the benchmark and transfer learning comparisons are affected by both feature timing and sample composition.

Figure 5.2 makes this limitation explicit by summarizing how the number of valid modelling cells changes across early-life feature weeks. The practical consequence is that the model evaluated at week 10 is not necessarily evaluated on exactly the same sample population as the model evaluated at week 5. Thus, a performance change across weeks can reflect both the information content of the later feature

and the changing set of cells that remain available for modelling.

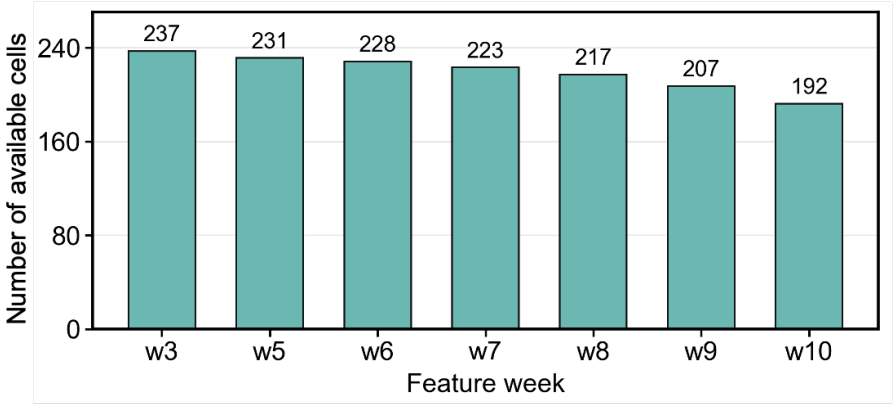


Figure 5.2: Availability of valid modelling cells across early-life feature weeks. The number of usable cells can change as the feature week increases because some cells may already have reached the end-of-life (EOL) threshold or may not have a valid feature observation at the selected week.

This limitation is important for interpreting Tables 4.5 and 5.1. The results show that fine-tuning improves performance when features are extracted at weeks 5, 6, 9, and 10, but worsens performance at weeks 7 and 8. This indicates that the transfer learning result is sensitive to the selected feature week. However, this sensitivity should not be interpreted simply as evidence that some weeks are intrinsically better than others. Since the available cell set may change across feature weeks, the observed performance difference can be affected by both the temporal maturity of the extracted features and the composition of the valid samples included in each experiment.

Therefore, these results should be interpreted as a sensitivity analysis rather than as a definitive optimization over feature week. A stronger conclusion would require controlling the valid cell set across all weeks, so that the same source, fine-tuning, and target-test cells are used in each comparison. Alternatively, the analysis could explicitly separate the effect of feature maturity from the effect of changing sample availability. This would make it possible to determine whether the performance variation is mainly caused by more informative later-life features, by changes in the evaluated cell population, or by the interaction between these two factors.

Table 5.1: Week sensitivity summary for the remaining useful life (RUL) transfer learning experiment. Positive Delta values indicate error reduction from fine-tuning relative to the benchmark.

Week	Benchmark MAE	Fine-tuned MAE	ΔMAE (%)	Benchmark MAPE (%)	Fine-tuned MAPE (%)	ΔMAPE (%)
w5	4.18	3.77	+9.8	14.62	13.67	+6.5
w6	3.77	3.74	+0.8	13.80	13.35	+3.3
w7	4.53	4.57	-1.0	14.81	15.76	-6.4
w8	4.07	4.54	-11.7	13.49	16.93	-25.5
w9	5.06	4.70	+7.3	16.86	15.35	+9.0
w10	4.59	3.77	+17.9	15.13	11.77	+22.2

For this reason, the week sensitivity result is best understood as a practical robustness check. It shows that the transfer learning workflow is affected by the amount and type of early-life information available, while also warning that week-based comparisons can be confounded by changes in which cells remain valid for modelling. This limitation should be considered when using early-week RUL prediction results to guide practical deployment.

5.5 Subjectivity of Domain Definition

The definition of source and target domains is not unique. In this thesis, domains are mainly defined by cell group and operating condition, which is a reasonable choice for studying cross-condition generalization. However, other domain definitions are possible. For example, domains could be defined by lifetime range, degradation pattern, material batch, calendar age, or combinations of operating variables. Each definition would create a different transfer learning problem and could change the apparent difficulty of the task.

This subjectivity matters because transfer learning is evaluated relative to the chosen source and target split. A low capacity or otherwise difficult target domain split can act as a stress test, revealing whether the model can adapt to target cells with degradation behaviour that is not well represented by the source data. Such a split may be valuable for evaluating robustness, but it may also increase the risk of negative transfer. In contrast, an easier split may produce stronger transfer gains while being less informative about difficult deployment scenarios.

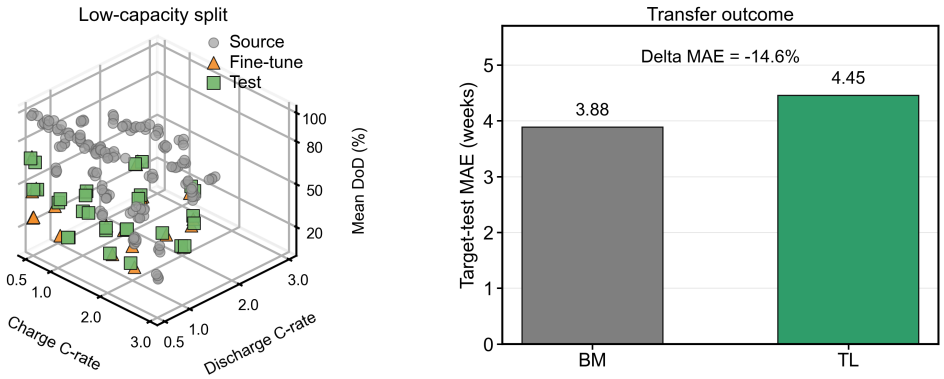


Figure 5.3: Operating condition distribution and target-test MAE comparison for the low capacity target split. The left panel shows source, target fine-tuning, and target-test cells in operating condition space, while the right panel compares the benchmark and transfer learning MAE. The higher transfer learning MAE indicates negative transfer under this domain definition.

Figure 5.3 illustrates this issue through a low capacity target split. The left panel shows the corresponding source, fine-tuning, and target-test cells in operating condition space, while the right panel compares the target-test MAE of the benchmark

and transfer learning models. In this split, fine-tuning increases the MAE from 3.88 to 4.45 weeks, as summarized in Table 5.2. This result is a useful counterexample to the main rerun400 improvement: transfer learning can become negative when the target fine-tuning cells do not provide sufficient support for the target-test distribution or when the domain definition introduces a biased adaptation task.

Table 5.2: Low capacity domain split result. Positive Delta MAE would indicate improvement from fine-tuning, whereas the negative value indicates worse performance after transfer learning.

Domain setting	Benchmark MAE	Fine-tuned MAE	ΔMAE (%)	Interpretation
Low capacity target split	3.88	4.45	-14.6	Transfer learning performs worse than the benchmark, suggesting negative transfer under this split.

The results therefore should not be interpreted as properties of transfer learning in isolation from domain design. They are properties of the transfer learning workflow under the domain definitions used in this thesis. Figures 3.4, 3.5, 5.1, and 5.3 together show that the target-test problem is shaped by operating condition separation, lifetime-distribution differences, and the particular way in which the source and target domains are defined. Future studies should report domain definitions explicitly and evaluate whether conclusions remain stable under alternative source and target partitions.

5.6 Influence of Fine-Tuning Cell Selection

The seed selection sensitivity results show that target domain fine-tuning cell selection has a strong influence on transfer performance. In Table 4.6, positive transfer is observed in 12 out of 21 selections, but some selections produce substantial negative transfer. The ΔMAE values range from -32.9% to $+16.2\%$, with a mean of only $+0.5\%$ across the seed sweep. This means that the main rerun400 improvement is real for that configuration, but it should be interpreted together with the broader variability across target domain splits.

Table 5.3: Seed selection sensitivity summary for target domain fine-tuning/test splitting. The statistics are calculated across 21 target domain selections using Delta MAE.

Statistic	ΔMAE (%)
Number of selections	21
Mean	+0.5
Standard deviation	11.9
Median	+1.6
Minimum	-32.9
Maximum	+16.2
Positive transfer cases	12 / 21

The reason is that the fine-tuning subset determines what target domain behaviour the model sees before final testing. If the selected fine-tuning cells are representative of the held-out target-test cells, fine-tuning can adapt the pretrained model in a useful direction. If they are not representative, the model may adapt toward a subset of target behaviour that does not match the test cells, leading to weaker improvement or negative transfer. This effect is especially important when the number of target fine-tuning cells is limited.

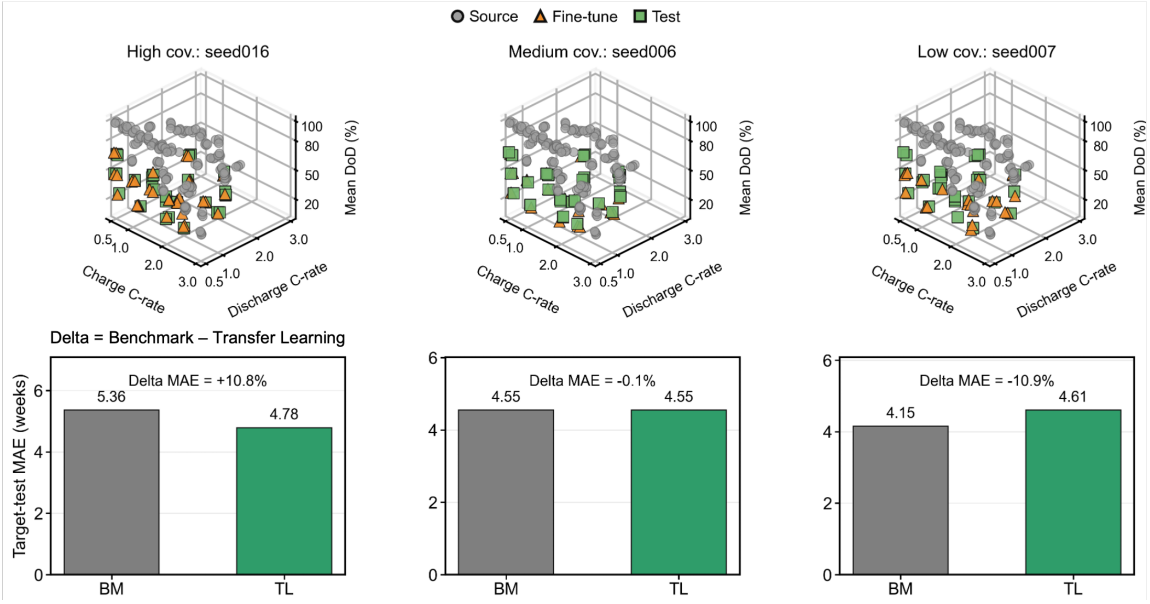


Figure 5.4: Representative target domain fine-tuning cell selection examples. The upper row shows target fine-tuning and target-test cells in operating condition space, and the lower row compares benchmark and transfer learning MAE for the corresponding selections. The examples illustrate that fine-tuning outcome can change substantially with the selected target cells.

Figure 5.4 compares three representative target fine-tuning cell selection cases. The upper panels visualize the selected fine-tuning cells and the remaining target-test cells in the operating condition space, while the lower panels report the corresponding MAE values of the benchmark and transfer learning models. These examples are chosen to represent high, medium, and low levels of operating condition coverage. In this context, coverage describes how well the selected fine-tuning cells represent the operating conditions of the target-test cells. Therefore, the figure provides an intuitive comparison between the spatial distribution of the fine-tuning cells and the resulting transfer learning performance.

The descriptive proximity metric reported in Table 5.4 is used only to summarize the operating condition closeness between the fine-tuning and test cells in these examples. It provides a compact way to describe the relative proximity shown in the upper panels of Figure 5.4. However, it is not introduced as a formal evaluation index. Therefore, in this section, the metric should be interpreted only as an auxiliary descriptive quantity for these representative cases, rather than as a general criterion for evaluating all fine-tuning selections.

Table 5.4: Representative target fine-tuning cell selection cases. The proximity values are descriptive operating condition distances used only for interpreting these examples.

Case	Fine-tuning proximity	Benchmark MAE	Fine-tuned MAE	ΔMAE (%)	Interpretation
High coverage	0	5.36	4.78	+10.8	Strong operating condition support; transfer improves.
Medium coverage	0.01	4.55	4.55	-0.1	Close operating conditions; transfer gain is negligible.
Low coverage	0.06	4.15	4.61	-10.9	Weaker target-test support; transfer degrades.

Overall, these examples support a cautious interpretation. Better coverage of the operating condition space can help make transfer more stable and can contribute to positive transfer, as in the high coverage case. However, operating condition coverage alone cannot fully explain every seed outcome, because degradation trajectory shape, lifetime range, cell-to-cell variation, and target-test uncertainty may also matter. Therefore, the results point toward the need for better target battery selection strategies rather than only more random seed repetitions. In this thesis, the proximity metric is used descriptively to support this discussion, not as a formal performance metric.

5.7 Conclusion

This thesis used the ISU-ILCC dataset to study cross-condition battery health diagnostics and prognostics. The dataset is suitable for this purpose because the cells are aged under diverse charge C-rates, discharge C-rates, and depth-of-discharge (DoD) levels. These operating conditions are physically related to battery ageing and therefore provide a meaningful basis for defining source and target domains. This enables the thesis to investigate whether models trained under one set of ageing conditions can generalize to cells aged under different conditions, and whether limited target-domain information can improve this generalization.

The SOH estimation results show that diagnostic health estimation is feasible under cross-condition splits when IC-based features are used. The selected IC-based representation contains useful degradation information, and the regression results confirm that the basic feature extraction and modelling pipeline is functional. However, the observed domain divergence in the SOH task is limited compared with the later RUL task. This is because SOH estimation is mainly a point-wise diagnostic problem, where the target variable changes relatively smoothly with degradation state. Therefore, SOH estimation is best interpreted as a preliminary diagnostic task. It verifies the validity of the feature and modelling pipeline and provides initial evidence of cross-condition effects, but it does not represent the main transfer learning challenge of the thesis.

The RUL prediction experiments show a clearer and more demanding cross-condition prognostics challenge. Compared with SOH estimation, RUL prediction requires the model to infer the remaining lifetime trajectory from limited early or

intermediate degradation information, which makes it more sensitive to domain divergence. In the representative target-test split, the fine-tuning model selected by Optuna with the TPE sampler improves prediction relative to the benchmark model, reducing MAE from 4.18 to 3.77 weeks. This result supports the use of source-domain pretraining followed by target-domain fine-tuning for cross-condition RUL prediction. At the same time, the broader sensitivity analysis shows that this improvement is conditional on feature week, domain definition, and target fine-tuning cell selection. The final conclusion is therefore that fine-tuning-based transfer learning is promising for cross-condition battery RUL prediction under limited target data, but its reliability depends on careful experimental design and on how well the selected target cells represent the final target-test task.

5.8 Future Work

Future work can be organized into three main directions. First, a systematic fine-tuning cell selection mechanism should be developed to improve target-domain coverage and transfer robustness. Instead of choosing target cells randomly, future studies could use coverage-aware selection, active learning, informative sampling, or uncertainty-based criteria to select cells that are expected to support adaptation most effectively. Such methods could consider operating conditions, early degradation trajectory shape, SOH/RUL distribution, prediction uncertainty, or diversity among candidate target cells.

Second, controlled feature-extraction week studies should be performed with consistent cell sets across observation weeks. In the present analysis, changing the feature week can also change which cells remain valid for modelling, making it difficult to separate the intrinsic informativeness of later features from dataset-size and sample-composition effects. A controlled study with fixed cell sets would allow the effect of feature timing to be evaluated more directly and would clarify whether later-week features improve prediction because they contain more degradation information or because the available sample population has changed.

Third, the workflow should be extended to dynamic cycling profiles and additional battery datasets for broader validation, because recent EV-representative ageing studies show that dynamic discharge profiles can lead to substantially different degradation behaviour from conventional constant-current protocols [4, 5]. The present thesis focuses on controlled constant-current cycling data from one dataset and one main RUL transfer workflow. Future work should test whether the same conclusions hold under dynamic load profiles, different chemistries, different cell formats, alternative ageing protocols, and independent datasets. Such validation would clarify whether the observed transfer-learning behaviour is dataset specific or reflects a more general pattern in battery RUL prediction.

Bibliography

[1] Vetter, J., Novák, P., Wagner, M. R., Veit, C., Möller, K.-C., Besenhard, J. O., Winter, M., Wohlfahrt-Mehrens, M., Vogler, C., and Hammouche, A. (2005) Ageing mechanisms in lithium-ion batteries. *Journal of Power Sources*, 147(1–2), 269–281.

[2] Birkl, C. R., Roberts, M. R., McTurk, E., Bruce, P. G., and Howey, D. A. (2017) Degradation diagnostics for lithium ion cells. *Journal of Power Sources*, 341, 373–386.

[3] Plett, G. L. (2015) *Battery Management Systems, Volume II: Equivalent-Circuit Methods*. Artech House, Norwood, MA.

[4] Keil, P. and Jossen, A. (2017) Impact of dynamic driving loads and regenerative braking on the aging of lithium-ion batteries in electric vehicles. *Journal of The Electrochemical Society*, 164(13), A3081–A3092.

[5] Geslin, A., Xu, L., Ganapathi, D., Moy, K., Chueh, W. C., and Onori, S. (2025) Dynamic cycling enhances battery lifetime. *Nature Energy*, 10, 172–180.

[6] Zhang, J. and Lee, J. (2011) A review on prognostics and health monitoring of Li-ion battery. *Journal of Power Sources*, 196(15), 6007–6014.

[7] Lipu, M. S. H., Hannan, M. A., Hussain, A., Hoque, M. M., Ker, P. J., Saad, M. H. M., and 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*, 205, 115–133.

[8] Berecibar, M., Gandiaga, I., Villarreal, I., Omar, N., Van Mierlo, J., and Van den Bossche, P. (2016) Critical review of state of health estimation methods of Li-ion batteries for real applications. *Renewable and Sustainable Energy Reviews*, 56, 572–587.

[9] Ge, M.-F., Liu, Y., Jiang, X., and Liu, J. (2021) A review on state of health estimations and remaining useful life prognostics of lithium-ion batteries. *Measurement*, 174, 109057.

[10] Zhao, J., Zhu, Y., Zhang, B., Liu, M., Wang, J., Liu, C., and Hao, X. (2023) Review of state estimation and remaining useful life prediction methods for lithium-ion batteries. *Sustainability*, 15(6), 5014.

[11] Nuhic, A., Terzimehic, T., Soczka-Guth, T., Buchholz, M., and Dietmayer, K. (2013) Health diagnosis and remaining useful life prognostics of lithium-ion batteries using data-driven methods. *Journal of Power Sources*, 239, 680–688.

[12] Ecker, M., Nieto, N., Käbitz, S., Schmalstieg, J., Blanke, H., Warnecke, A., and Sauer, D. U. (2014) Calendar and cycle life study of Li(NiMnCo)O_2-based 18650 lithium-ion batteries. *Journal of Power Sources*, 248, 839–851.

[13] Jin, S., Sui, X., Huang, X., Wang, S., Teodorescu, R., and Stroe, D.-I. (2021) Overview of machine learning methods for lithium-ion battery remaining useful lifetime prediction. *Electronics*, 10(24), 3126.

[14] Meng, H. and Li, Y. (2019) A review on prognostics and health management (PHM) methods of lithium-ion batteries. *Renewable and Sustainable Energy Reviews*, 116, 109405.

[15] 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., and Braatz, R. D. (2019) Data-driven prediction of battery cycle life before capacity degradation. *Nature Energy*, 4, 383–391.

[16] 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., and Chueh, W. C. (2020) Closed-loop optimization of fast-charging protocols for batteries with machine learning. *Nature*, 578, 397–402.

[17] Che, Y., Hu, X., Lin, X., Guo, J., and Teodorescu, R. (2023) Health prognostics for lithium-ion batteries: mechanisms, methods, and prospects. *Energy & Environmental Science*, 16, 338–371.

[18] Li, T., Zhou, Z., Thelen, A., Howey, D. A., and Hu, C. (2024) Predicting battery lifetime under varying usage conditions from early aging data. *Cell Reports Physical Science*, 5(4), 101891.

[19] Widodo, A., Shim, M.-C., Caesarendra, W., and Yang, B.-S. (2011) Intelligent prognostics for battery health monitoring based on sample entropy. *Expert Systems with Applications*, 38(9), 11763–11769.

[20] André, D., Nuhic, A., Soczka-Guth, T., and Sauer, D. U. (2013) Comparative study of a structured neural network and an extended Kalman filter for state of health determination of lithium-ion batteries in hybrid electric vehicles. *Engineering Applications of Artificial Intelligence*, 26(3), 951–961.

[21] Williard, N., He, W., Osterman, M., and Pecht, M. (2013) Comparative analysis of features for determining state of health in lithium-ion batteries. *International Journal of Prognostics and Health Management*, 4(1).

[22] Richardson, R. R., Osborne, M. A., and Howey, D. A. (2017) Gaussian process regression for forecasting battery state of health. *Journal of Power Sources*, 357, 209–219.

[23] Dubarry, M., Truchot, C., and Liaw, B. Y. (2012) Synthesize battery degradation modes via a diagnostic and prognostic model. *Journal of Power Sources*, 219, 204–216.

[24] Zheng, L., Zhu, J., Lu, D. D.-C., Wang, G., and He, T. (2018) Incremental capacity analysis and differential voltage analysis based state of charge and capacity estimation for lithium-ion batteries. *Energy*, 150, 759–769.

[25] Krupp, A., Ferg, E., Schuldt, F., Derendorf, K., and Agert, C. (2021) Incremental Capacity Analysis as a State of Health Estimation Method for Lithium-Ion Battery Modules with Series-Connected Cells. *Batteries*, 7(1), 2.

[26] Huang, K., Yao, K., Guo, Y., and Lv, Z. (2023) State of health estimation of lithium-ion batteries based on fine-tuning or rebuilding transfer learning strategies combined with new features mining. *Energy*, 282, 128739.

[27] Fly, A. and Chen, R. (2020) Rate dependency of incremental capacity analysis (dQ/dV) as a diagnostic tool for lithium-ion batteries. *Journal of Energy Storage*, 29, 101329.

[28] Pan, S. J. and Yang, Q. (2010) A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10), 1345–1359.

[29] Weiss, K., Khoshgoftaar, T. M., and Wang, D. (2016) A survey of transfer learning. *Journal of Big Data*, 3, 9.

[30] Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., Marchand, M., and Lempitsky, V. (2016) Domain-adversarial training of neural networks. *Journal of Machine Learning Research*, 17(59), 1–35.

[31] Li, Y., Sheng, H., Cheng, Y., Stroe, D.-I., and Teodorescu, R. (2020) State-of-health estimation of lithium-ion batteries based on semi-supervised transfer component analysis. *Applied Energy*, 277, 115504.

[32] Amogne, Z. E., Wang, F.-K., and Chou, J.-H. (2023) Transfer Learning Based on Transferability Measures for State of Health Prediction of Lithium-Ion Batteries. *Batteries*, 9(5), 280.

[33] Shen, L., Li, J., Meng, L., Zhu, L., and Shen, H. T. (2024) Transfer Learning-Based State of Charge and State of Health Estimation for Li-Ion Batteries: A Review. *IEEE Transactions on Transportation Electrification*, 10(1), 1465–1481.

[34] Chou, J.-H., Wang, F.-K., and Lo, S.-C. (2023) A Novel Fine-Tuning Model Based on Transfer Learning for Future Capacity Prediction of Lithium-Ion Batteries. *Batteries*, 9(6), 325.

[35] Gu, B. and Liu, Z. (2024) Transfer Learning-Based Remaining Useful Life Prediction Method for Lithium-Ion Batteries Considering Individual Differences. *Applied Sciences*, 14(2), 698.

[36] Bergstra, J. and Bengio, Y. (2012) Random Search for Hyper-Parameter Optimization. *Journal of Machine Learning Research*, 13, 281–305.

[37] Snoek, J., Larochelle, H., and Adams, R. P. (2012) Practical Bayesian Optimization of Machine Learning Algorithms. In *Advances in Neural Information Processing Systems 25*, 2951–2959.

[38] Akiba, T., Sano, S., Yanase, T., Ohta, T., and Koyama, M. (2019) Optuna: A Next-Generation Hyperparameter Optimization Framework. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2623–2631.

[39] Bergstra, J. S., Bardenet, R., Bengio, Y., and Kégl, B. (2011) Algorithms for Hyper-Parameter Optimization. In *Advances in Neural Information Processing Systems 24*, 2546–2554.

[40] Iowa State University DataShare (2024) ISU-ILCC NMC/Gr battery aging dataset. Available at DOI: `10.25380/iastate.22582234.44`.

[41] Hyndman, R. J. and Koehler, A. B. (2006) Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4), 679–688.

A

Appendix

A.1 Dataset and Cell Availability

A.1.1 Dataset Summary

Table A.1 lists the valid-cell, group, label-availability, and retained-cell counts used for RUL modelling.

Table A.1: Dataset summary under the end-of-life (EOL) threshold defined at 70% state of health (SOH).

Item	Value	Source
Valid cells listed before feature-table construction for the EOL threshold defined at 70% SOH	251	`Valid_cells.csv`
Operating-condition groups	63	`Groupcondi.csv`
Cells with available lifetime label under the EOL threshold defined at 70% SOH	247	`ivas_lifetime_eol_availability.csv`
Cells missing lifetime label under the EOL threshold defined at 70% SOH	4	`ivas_lifetime_eol_availability.csv`
Cells retained in final multi-week feature table for the EOL threshold defined at 70% SOH	242	final feature table for the EOL threshold defined at 70% SOH
Groups retained in final multi-week feature table for the EOL threshold defined at 70% SOH	63	final feature table for the EOL threshold defined at 70% SOH
Valid-list cells not retained in final feature table for the EOL threshold defined at 70% SOH	9	comparison between `Valid_cells.csv` and final feature table

A.1.2 Label Availability at 70% SOH

At the 70% SOH EOL threshold, 247 of 251 cells have lifetime labels. Figure A.1 shows label availability across EOL thresholds.

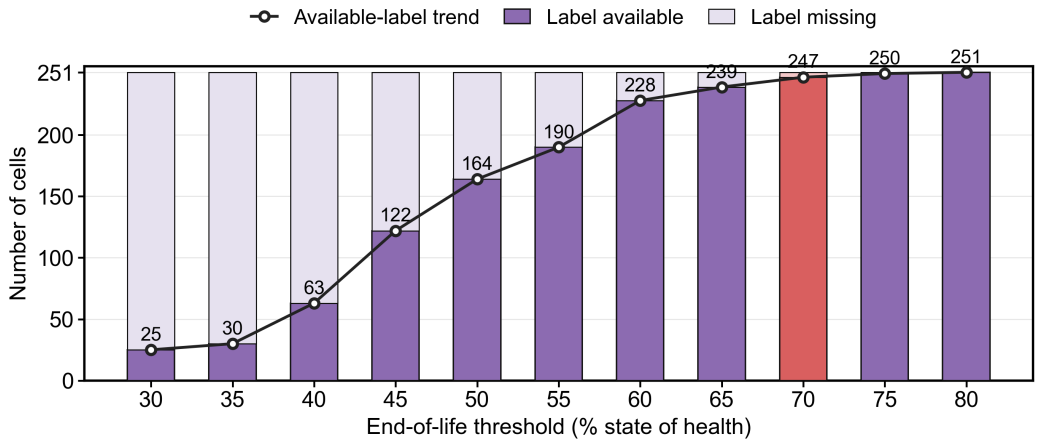


Figure A.1: Label availability across end-of-life (EOL) thresholds. Purple bars indicate available labels, light purple bars indicate missing labels, and the black line shows the available-label trend. The 70% SOH threshold is highlighted.

A.1.3 Week-Based Feature Availability

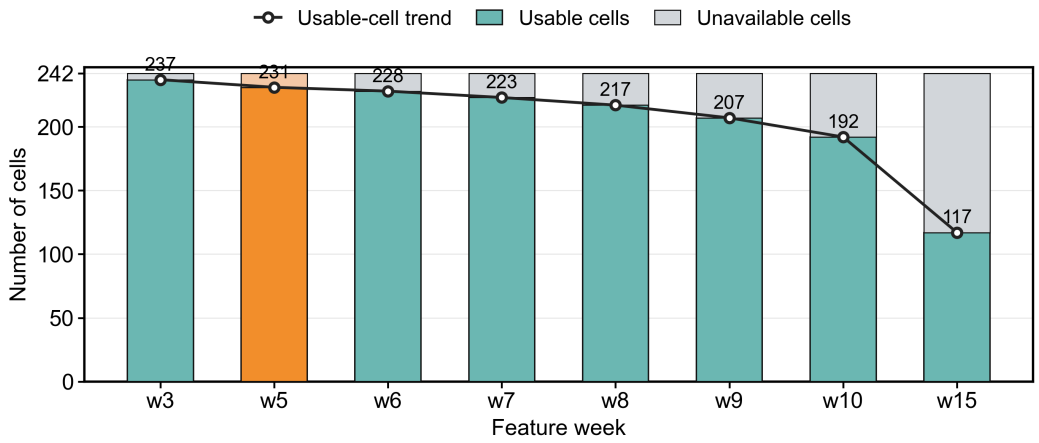


Figure A.2: Cell availability across feature-extraction weeks. Coloured bars show usable cells, grey bars show unusable cells, and the black line shows the usable-cell trend. Week 5 is highlighted.

Table A.2 reports feature-week availability. Usable non-NaN cells decrease from 237 at week 3 to 192 at week 10 and 117 at week 15.

Table A.2: Week-based feature availability in the final multi-week feature table.

Week label	Usable non-NaN cells	Status-ok cells	Feature/NaN unusable cells	Total cells
w3	237	240	5	242
w5	231	236	11	242
w6	228	232	14	242

Table A.2: Week-based feature availability in the final multi-week feature table (continued).

Week label	Usable non-NaN cells	Status-ok cells	Feature/NaN unusable cells	Total cells
w7	223	228	19	242
w8	217	224	25	242
w9	207	221	35	242
w10	192	212	50	242
w15	117	148	125	242

A.1.4 Cell-Retention Summary

Table A.3 maps the valid-cell list to final feature-table retention status.

Table A.3: Cell-retention summary for the final feature table under the end-of-life (EOL) threshold defined at 70% state of health (SOH).

Availability status	Cell count	Fraction of valid cells	Example cells
Retained in final feature table under the EOL threshold defined at 70% SOH	242	0.964143	G1C1, G1C4, G2C1, G2C2, G2C3, G2C4, G3C1, G3C2
Not retained in final feature table, reason not encoded	5	0.0199203	G1C2, G1C3, G6C3, G18C1, G26C3
Missing label and week features under the EOL threshold defined at 70% SOH	1	0.00398406	G14C4
Missing label under the EOL threshold defined at 70% SOH	3	0.0119522	G57C1, G57C2, G57C4

The modelling dataset changes with feature-extraction week; week-based RUL results should therefore be interpreted with cell retention and sample composition.

A.2 Engineered Feature Definitions

A.2.1 Notation and Feature Categories

Let w_0 denote the baseline diagnostic week and w_k a later early-life diagnostic week; in the main experiment, $k = 5$. For any descriptor x:

$$\Delta x_{w_k - w_0} = x_{w_k} - x_{w_0}. \tag{A.1}$$

Here, dQ/dV is the incremental-capacity curve, CV time is the constant-voltage charging time, C_{chg} is the charge C-rate, and DoD is depth of discharge.

The features are grouped as IC-curve, CV-time, usage-condition, and capacity-response descriptors. IC-curve features use selected dQ/dV voltage windows; CV-time features use constant-voltage charging behaviour; usage-condition features use DoD, C_{chg}, and their interaction; capacity-response features use ΔQ^1 and ΔQ^3.

A.2.2 Feature Definitions

The suffix `w5` indicates early-life relative to baseline w_0, except `f10_w5`, which uses the initial CV time at w_0.

1. `f1_w5` (IC-curve feature)

   $\log(|\mathrm{mean}(\Delta(dQ/dV)_{w_5-w_0}^{3.6-3.9\,\mathrm{V}})|)$

   Average early-life change in the incremental-capacity curve within the 3.6–3.9 V window.

2. `f2_w5` (CV-time feature)

   $\log(|\Delta\text{CV time}_{w_5-w_0}|)$

   Absolute early-life change in constant-voltage charging time.

3. `f3_w5` (Usage-condition feature)

   DoD

   Depth of discharge, representing the fraction of available capacity used during cycling.

4. `f4_w5` (Capacity-response feature)

   $\Delta Q^{1}_{w_5-w_0}$

   Early-life change in the first capacity-related descriptor.

5. `f5_w5` (Usage-condition feature)

   $C_{\mathrm{chg}}^{0.5}\mathrm{DoD}^{0.5}$

   Interaction between charge C-rate and depth of discharge.

6. `f6_w5` (Usage-condition feature)

   C_{chg}

   Charge C-rate, describing charging current relative to nominal cell capacity.

7. `f7_w5` (IC-curve feature)

   $\log(\mathrm{var}(\Delta(dQ/dV)_{w_5-w_0}^{3.0-3.6\,\mathrm{V}}))$

   Variance of early-life incremental-capacity curve change within the 3.0–3.6 V window.

8. `f8_w5` (Capacity-response feature)

   $\Delta Q^{3}_{w_5-w_0}$

   Early-life change in the third capacity-related descriptor, complementary to the first descriptor.

9. `f9_w5` (IC-curve feature)

   $\log(|\mathrm{mean}(\Delta(dQ/dV)_{w_5-w_0}^{3.0-3.6\,\mathrm{V}})|)$

   Average early-life change in the incremental-capacity curve within the lower-voltage 3.0–3.6 V window.

10. `f10_w5` (CV-time feature)

    $\log(|\text{CV time}_{w_0}|)$

    Initial constant-voltage charging time before the diagnostic ageing window.

A.2.3 Early-life Feature Value Summary

Table A.4 gives early-life feature distributions and lifetime-label correlations.

Table A.4: Early-life feature value summary and feature-lifetime correlations under the end-of-life (EOL) threshold defined at 70% state of health (SOH).

Feature	n non-NaN	Mean	Std.	Min.	Median	Max.	Pearson r	Spearman r
f1_w5	236	-2.8026	0.3153	-3.6162	-2.7476	-1.4751	-0.8278	-0.8876
f2_w5	236	5.5720	0.6503	2.3979	5.7071	8.1131	0.2857	0.4490
f3_w5	242	0.6298	0.2622	0.0346	0.6164	0.9965	-0.7012	-0.7152
f4_w5	231	-0.0092	0.0177	-0.0775	-0.0043	0.0279	0.2159	0.2180
f5_w5	242	0.9545	0.2909	0.2202	0.9636	1.5560	-0.7088	-0.7218
f6_w5	242	1.6217	0.6829	0.5000	1.6875	3.0000	-0.2118	-0.2206
f7_w5	236	-6.7651	0.7613	-8.4527	-6.8125	-4.1603	-0.4256	-0.4487
f8_w5	231	0.0180	0.0121	0.0021	0.0155	0.0832	-0.4080	-0.4835
f9_w5	236	-4.6251	1.0739	-11.3856	-4.4304	-2.4982	-0.1652	-0.2580
f10_w5	242	6.9559	0.1307	6.6958	6.9527	8.3687	0.2072	0.2946

A.2.4 Early-life Feature-Lifetime Correlations

Table A.5 reports feature-lifetime Pearson and Spearman correlations. The largest absolute correlations are for `f1_w5`, `f3_w5`, and `f5_w5`.

Table A.5: Feature-lifetime correlations for the early-life remaining useful life (RUL) feature set under the end-of-life (EOL) threshold defined at 70% state of health (SOH).

Feature	Feature description	Pearson r	Spearman r	n
f1_w5	log|mean dQ/dV delta 3.6–3.9 V| [w5]	-0.828	-0.888	241
f2_w5	log|delta CV time| [w5]	0.286	0.449	241
f3_w5	DoD [w5]	-0.701	-0.715	247
f4_w5	delta Q1 DVA [w5]	0.216	0.218	236
f5_w5	sqrt(Cchg)*sqrt(DoD) [w5]	-0.709	-0.722	247
f6_w5	Cchg [w5]	-0.212	-0.221	247
f7_w5	log|var dQ/dV delta 3.0–3.6 V| [w5]	-0.426	-0.449	241
f8_w5	delta Q3 DVA [w5]	-0.408	-0.483	236
f9_w5	log|mean dQ/dV delta 3.0–3.6 V| [w5]	-0.165	-0.258	241
f10_w5	log|CV time w0| [w5]	0.207	0.295	247

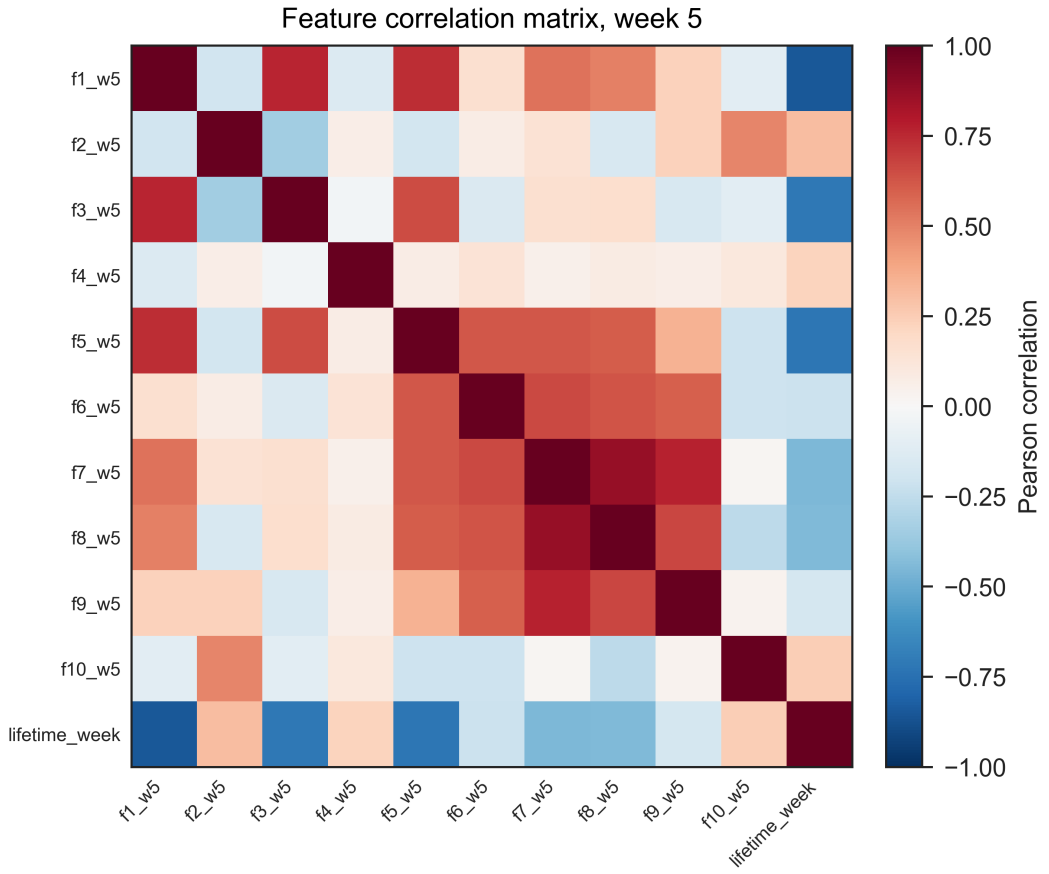


Figure A.3: Feature correlation matrix for the early-life remaining useful life (RUL) feature set under the 70% SOH EOL threshold.

These definitions and statistics provide the feature context for the early-life RUL experiments.

A.3 Hyperparameter Search and Final Model Configuration

A.3.1 Search Procedure

Stage 1 selected the source pretraining feature subset, MLP architecture, dropout, activation, epochs, learning rate, and weight decay. The selected source model initialized transfer learning.

Stage 2 selected the fine-tuning learning rate, weight decay, epochs, frozen hidden layers, target support ratio, and source replay weight using target-domain validation MAE.

Table A.6 summarizes the search components for the main early-life RUL experiment.

Table A.6: Hyperparameter search components for the main early-life remaining useful life (RUL) transfer learning experiment.

Search stage	Configuration components	Role in workflow
Stage 1 source search	Features, hidden dimensions, dropout, activation, epochs, learning rate, weight decay	Selects the pretrained source model used as transfer initialization.
Stage 2 fine-tuning search	Fine-tuning learning rate, fine-tuning weight decay, fine-tuning epochs, frozen hidden layers, target support ratio, source replay weight	Selects the target-domain adaptation strategy applied to the pretrained model.
Final evaluation	Feature week and end-of-life threshold definition	Fixes the final test protocol used for the reported main experiment.

A.3.2 Selected Stage 2 Fine-Tuning Configuration

Table A.7: Selected Stage 2 fine-tuning configuration for the main early-life remaining useful life (RUL) experiment.

Parameter	Selected value
ft_lr	0.0002703005473042
ft_weight_decay	3.380298296224284e-06
ft_epochs	400
ft_freeze_hidden_layers	3
target_support_ratio	0.67
transfer_replay_weight	1.0
target_ft_val_mae	1.6633552307128905

A.3.3 Final Selected Configuration

Table A.8: Final selected model configuration for the main early-life remaining useful life (RUL) experiment under the end-of-life (EOL) threshold defined at 70% state of health (SOH).

Section	Parameter	Value
source pretraining	features	`f1_w5,f6_w5`
source pretraining	hidden_dims	`64,64,64,64`
source pretraining	dropout	0.0
source pretraining	activation	relu
source pretraining	epochs	800
source pretraining	lr	0.0009741957660925079
source pretraining	weight_decay	0.00013369738668013456
source pretraining	source_stage1_val_mae	1.794578163840554
target fine-tuning	ft_lr	0.0002703005473042
target fine-tuning	ft_weight_decay	3.380298296224284e-06
target fine-tuning	ft_epochs	400
target fine-tuning	ft_freeze_hidden_layers	3
target fine-tuning	target_support_ratio	0.67
target fine-tuning	transfer_replay_weight	1.0
target fine-tuning	target_ft_val_mae	1.6633552307128905

Table A.8: Final selected model configuration for the main early-life remaining useful life (RUL) experiment under the end-of-life (EOL) threshold defined at 70% state of health (SOH) (continued).

Section	Parameter	Value
final evaluation	feature_week	`w5`
final evaluation	EOL threshold	the end-of-life (EOL) threshold defined at 70% state of health (SOH)

The final evaluation uses early-life features and the EOL threshold defined at 70% SOH.

A.4 Sensitivity Check Results

A.4.1 Week-Based Sensitivity

Table A.9 compares benchmark, source-only, and fine-tuned transfer models for feature weeks 5–10. Target-test cells and groups vary by week, so these results should be read with Appendix A.1.

Table A.9: Week-level RUL metrics under the 70% SOH EOL threshold.

Model	Week	MAE	RMSE	R^2	MAPE (%)	sMAPE (%)	wMAPE (%)	n cells / groups
benchmark	5	4.181	5.421	0.257	14.621	15.909	15.522	29 / 16
source-only	5	5.104	6.491	-0.065	17.408	19.467	18.947	29 / 16
fine-tuned transfer	5	3.773	4.993	0.370	13.666	14.109	14.004	29 / 16
benchmark	6	3.769	5.106	0.332	13.804	13.966	13.904	30 / 17
source-only	6	3.896	5.150	0.321	13.671	14.588	14.372	30 / 17
fine-tuned transfer	6	3.739	4.910	0.383	13.351	13.913	13.793	30 / 17
benchmark	7	4.526	5.956	0.315	14.811	15.709	16.287	31 / 17
source-only	7	5.548	7.179	0.005	18.146	20.468	19.964	31 / 17
fine-tuned transfer	7	4.572	5.826	0.345	15.762	16.922	16.452	31 / 17
benchmark	8	4.065	5.647	0.384	13.492	14.341	14.628	31 / 17
source-only	8	4.237	5.937	0.319	13.644	14.536	15.245	31 / 17
fine-tuned transfer	8	4.540	5.887	0.331	16.934	15.682	16.336	31 / 17
benchmark	9	5.064	8.679	-0.253	16.862	15.363	17.705	33 / 17
source-only	9	5.079	7.176	0.143	15.788	16.770	17.757	33 / 17
fine-tuned transfer	9	4.696	7.793	-0.010	15.347	14.481	16.417	33 / 17
benchmark	10	4.591	7.826	-0.080	15.127	13.924	16.039	35 / 17
source-only	10	3.957	5.711	0.425	12.100	13.108	13.822	35 / 17
fine-tuned transfer	10	3.768	5.347	0.496	11.769	12.461	13.163	35 / 17

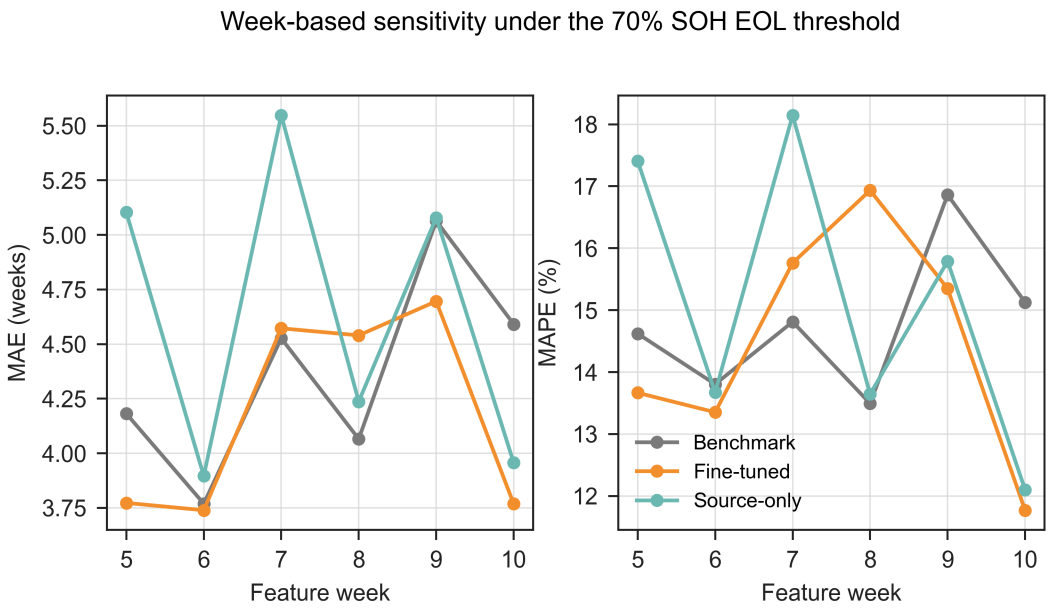


Figure A.4: Week-based sensitivity metrics for benchmark, source-only, and fine-tuned transfer models.

A.4.2 Random Target-Cell Selection Sensitivity

Table A.10 reports seed-level benchmark and fine-tuned transfer metrics for different target-cell splits. Positive MAE improvement means fine-tuning reduced benchmark MAE; negative improvement means it increased error.

Table A.10: Random target-cell selection sensitivity benchmark metrics under the 70% SOH EOL threshold.

Seed	n	**Benchmark MAE**	**Benchmark RMSE**	**Benchmark MAPE**	**Benchmark** R^2
seed000	29	5.162	6.579	18.147	-0.124
seed001	29	4.623	6.129	17.175	0.443
seed002	29	5.460	7.121	17.521	0.265
seed003	29	4.275	5.689	14.620	0.121
seed004	29	5.236	6.484	20.436	0.196
seed005	29	4.272	5.699	15.976	0.479
seed006	29	4.547	5.772	17.622	0.528
seed007	29	4.154	5.747	15.690	0.476
seed008	29	4.242	5.413	17.293	0.327
seed009	29	5.088	6.327	19.531	0.078
seed010	29	5.087	6.412	19.286	0.453
seed011	29	4.670	5.763	18.567	0.320
seed012	29	4.871	6.327	16.916	0.223
seed013	29	4.161	5.433	15.345	0.201
seed014	29	4.637	6.087	16.642	0.548
seed015	29	5.108	6.721	17.883	0.474
seed016	29	5.363	6.819	21.444	0.153
seed017	29	3.638	5.016	13.721	0.348
seed018	29	5.994	8.941	21.616	-0.424

Table A.10: Random target-cell selection sensitivity benchmark metrics under the 70% SOH EOL threshold (continued).

Seed	n	Benchmark MAE	Benchmark RMSE	Benchmark MAPE	Benchmark R^2
seed019	29	4.787	6.360	15.696	0.418

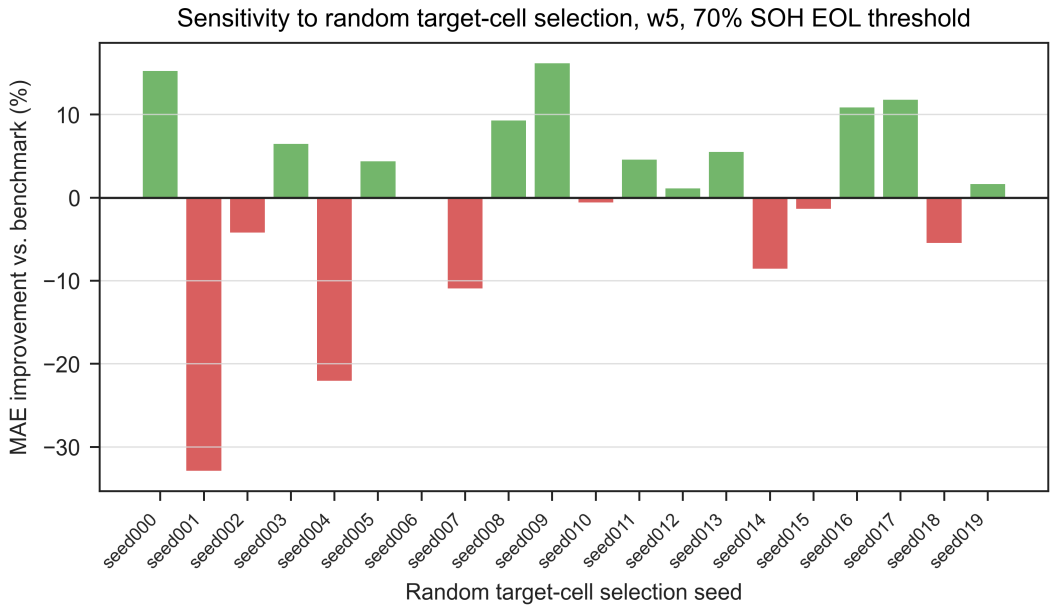


Figure A.5: Seed-level MAE improvement for random target-cell selection sensitivity.

Table A.11: Random target-cell selection sensitivity fine-tuned and improvement metrics under the 70% SOH EOL threshold.

Seed	n	Fine-tuned MAE	Fine-tuned RMSE	Fine-tuned MAPE	Fine-tuned R^2	ΔMAE (%)	ΔMAPE (%) / ΔR^2
seed000	29	4.376	5.720	15.814	0.150	15.235	12.855 / 0.274
seed001	29	6.142	10.382	23.516	-0.597	-32.861	-36.924 / -1.040
seed002	29	5.691	7.587	17.803	0.166	-4.219	-1.612 / -0.099
seed003	29	3.998	5.405	13.632	0.207	6.465	6.760 / 0.086
seed004	29	6.389	7.540	26.606	-0.086	-22.019	-30.194 / -0.283
seed005	29	4.086	5.301	16.719	0.549	4.359	-4.653 / 0.070
seed006	29	4.553	5.480	18.768	0.575	-0.131	-6.508 / 0.046
seed007	29	4.608	6.225	16.314	0.386	-10.932	-3.983 / -0.091
seed008	29	3.849	4.726	15.193	0.487	9.272	12.142 / 0.160
seed009	29	4.266	5.478	16.761	0.309	16.167	14.181 / 0.231
seed010	29	5.117	6.357	20.074	0.463	-0.585	-4.088 / 0.009
seed011	29	4.457	6.265	15.921	0.197	4.563	14.254 / -0.124
seed012	29	4.816	6.263	16.502	0.239	1.115	2.450 / 0.016
seed013	29	3.931	4.861	14.516	0.361	5.520	5.406 / 0.159
seed014	29	5.034	6.514	17.348	0.483	-8.548	-4.241 / -0.066
seed015	29	5.176	6.582	19.207	0.496	-1.330	-7.406 / 0.022
seed016	29	4.781	6.331	18.970	0.270	10.844	11.534 / 0.117

Table A.11: Random target-cell selection sensitivity fine-tuned and improvement metrics under the 70% SOH EOL threshold (continued).

Seed	n	Fine-tuned MAE	Fine-tuned RMSE	Fine-tuned MAPE	Fine-tuned R^2	ΔMAE (%)	ΔMAPE (%) / ΔR^2
seed017	29	3.209	4.001	12.605	0.586	11.775	8.138 / 0.237
seed018	29	6.320	10.615	22.860	-1.008	-5.450	-5.755 / -0.584
seed019	29	4.709	6.038	15.916	0.475	1.639	-1.399 / 0.057

Across 20 random splits, 10 improve MAE and 10 degrade it. The best case is `seed009` (+16.167% MAE improvement), and the worst is `seed001` (−32.861%).

The sensitivity checks show dependence on both feature-extraction week and target-cell selection.

B

Codes

B.1 Source Training Stage

Fits source-domain standardization and trains the source MLP.

Listing B.1: Core source-domain training and checkpoint export logic.

```
mu, sd = week_runner.ridge_utils.standardize_fit(
    source_train_df[list(x_cols)].to_numpy(dtype=float)
)

source_stage = mlp_base.train_stage(
    train_df=source_train_df,
    x_cols=x_cols,
    y_col=args.y_col,
    mu=mu,
    sd=sd,
    hidden_dims=hidden_dims,
    dropout=float(args.dropout),
    activation=args.activation,
    epochs=int(args.epochs),
    lr=float(args.lr),
    weight_decay=float(args.weight_decay),
    batch_size=int(args.batch_size),
    val_cell_frac=float(args.val_cell_frac),
    early_stop_patience=int(args.early_stop_patience),
    seed=int(args.seed),
    device=device,
    freeze_hidden_layers=0,
    stage_name="source_train",
    require_validation=True,
)

save_checkpoint(
    out_root / "source_checkpoint.pt",
    {
        "meta": {
            "features": list(feature_aliases),
            "x_cols": list(x_cols),
            "y_col": args.y_col,
            "hidden_dims": list(hidden_dims),
            "dropout": float(args.dropout),
            "activation": args.activation,
            "mu": np.asarray(mu, dtype=float).tolist(),
            "sd": np.asarray(sd, dtype=float).tolist(),
        },
        "source_metrics": source_metrics,
        "state_dict": source_stage["state_dict"],
    },
)
```

B.2 Target Fine-Tuning Stage

Initializes from the source checkpoint and fine-tunes on target support cells with optional source replay and layer freezing.

Listing B.2: Core target-domain fine-tuning and checkpoint export logic.

```
active_stage = train_stage_with_replay(
    train_df=target_ft_df,
    replay_df=source_train_df,
    replay_loss_weight=float(args.transfer_replay_weight),
    x_cols=x_cols,
    y_col=args.y_col,
    mu=np.asarray(mu, dtype=float),
    sd=np.asarray(sd, dtype=float),
    hidden_dims=hidden_dims,
    dropout=float(args.dropout),
    activation=args.activation,
    epochs=int(args.ft_epochs),
    lr=float(args.ft_lr),
    weight_decay=float(args.ft_weight_decay),
    batch_size=int(args.batch_size),
    val_cell_frac=float(args.val_cell_frac),
    seed=int(args.seed) + 1,
    device=device,
    init_state_dict=source_stage["state_dict"],
    freeze_hidden_layers=int(args.ft_freeze_hidden_layers),
    min_val_cells=int(args.min_target_val_cells),
    stage_name="target_finetune",
    ft_batch_mode=str(args.ft_batch_mode),
    ft_selection_mode=str(args.ft_selection_mode),
    ft_l2sp_weight=float(args.ft_l2sp_weight),
)

save_checkpoint(
    out_root / "finetuned_checkpoint.pt",
    {
        "meta": {
            "source_checkpoint": str(source_checkpoint_out),
            "features": list(feature_aliases),
            "x_cols": list(x_cols),
            "y_col": args.y_col,
            "hidden_dims": list(hidden_dims),
            "mu": np.asarray(mu, dtype=float).tolist(),
            "sd": np.asarray(sd, dtype=float).tolist(),
        },
        "trial_summary": trial_summary,
        "state_dict": active_stage["state_dict"],
    },
)
```

Listing B.3 shows the target loss with optional source replay and L2-SP terms.

Listing B.3: Core optimization loop with optional source replay and L2-SP regularization.

```
for xb, yb in train_loader:
    optimizer.zero_grad(set_to_none=True)
    pred = model(xb)
    target_loss = loss_fn(pred, yb)
    total_loss = target_loss

    if use_replay and replay_iter is not None:
        rb, ry = next(replay_iter)
        replay_loss = loss_fn(model(rb), ry)
        total_loss = total_loss + float(replay_loss_weight) * replay_loss

    if float(ft_l2sp_weight) > 0.0:
```

```
        l2sp_loss = sum(
            torch.sum((param - source_params[name]) ** 2)
            for name, param in model.named_parameters()
            if param.requires_grad and name in source_params
        ) / float(l2sp_numel)
        total_loss = total_loss + float(ft_l2sp_weight) * l2sp_loss

    total_loss.backward()
    optimizer.step()
```

B.3 Target Test Stage

Evaluates the fine-tuned model on held-out target-test cells and exports predictions and metrics.

Listing B.4: Target-test prediction export and metric summarization.

```
if args.evaluate_target_test:
    target_test_pred_df = week_runner.ridge_utils.build_prediction_df(
        target_test_df,
        args.y_col,
        mlp_base.predict_with_model(
            active_stage["model"],
            target_test_df,
            x_cols,
            np.asarray(mu, dtype=float),
            np.asarray(sd, dtype=float),
            device,
            int(args.batch_size),
        ),
        "target_test_finetuned",
    )

    target_test_overall = week_runner.ridge_utils.summarize_overall(
        target_test_pred_df,
        "target_test_finetuned",
        float(args.tail_q),
    )

    trial_summary.update(
        to_metrics(target_test_overall.iloc[0].to_dict(), "target_test")
    )
    target_test_pred_df.to_csv(out_root / "predictions_target_test.csv", index=False)
    target_test_overall.to_csv(out_root / "target_test_overall_metrics.csv", index=False)
```

Assembles the benchmark and transfer-learning report.

Listing B.5: Final benchmark and transfer-learning report assembly.

```
transfer_test = read_single_row_csv(
    out_dir / "transfer_model" / "test_overall_metrics.csv"
)
benchmark_test = read_single_row_csv(
    out_dir / "benchmark" / "test_overall_metrics.csv"
)

summary_report = {
    "benchmark_test_overall": benchmark_test,
    "transfer_test_overall": transfer_test,
    "selected_stage2_row": best_row,
    "selected_stage1_summary": source_summary,
    "final_package_root": str(out_dir),
}

save_json(out_dir / "stage3_final_report.json", summary_report)
```

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

CHALMERS
UNIVERSITY OF TECHNOLOGY