PhD Work Summary at Tsinghua & Berkeley

Machine Learning-Assisted Sustainable Remanufacturing, Reusing and Recycling for Lithium-ion Batteries

Dr. Shengyu Tao
Institute of Data and Information, Tsinghua University

Acknowledgments to Dr. Xuan Zhang, Dr. Guangmin Zhou and Dr. Scott Moura

2025/9/16

Chapter 1: Introduction

☐ Bio

☐ Some facts about me

Dr. Shengyu Tao

2023.9-2024.4	Tencent AI Lab	- Generative modeling for batteries - Conditional variational autoencoder development
2024.4-2024.9	Microsoft Research	- Physics-informed battery data generation - PyBaMM + AI -> Searching for parameters
2024.9-2025.9	Berkeley UNIVERSITY OF CALIFORNIA	- AI for battery lifecycle management - Data Science in battery management - Remanufacturing, reusing and recycling - PhD degree in Electrical Engineering
2022.9-2025.9	清華大學 Tsinghua University	

Table of Contents

Chapter 1: Introduction

Chapter 1: Introduction

☐ Background and Motivation

☐ Resource underutilization and environmental impact

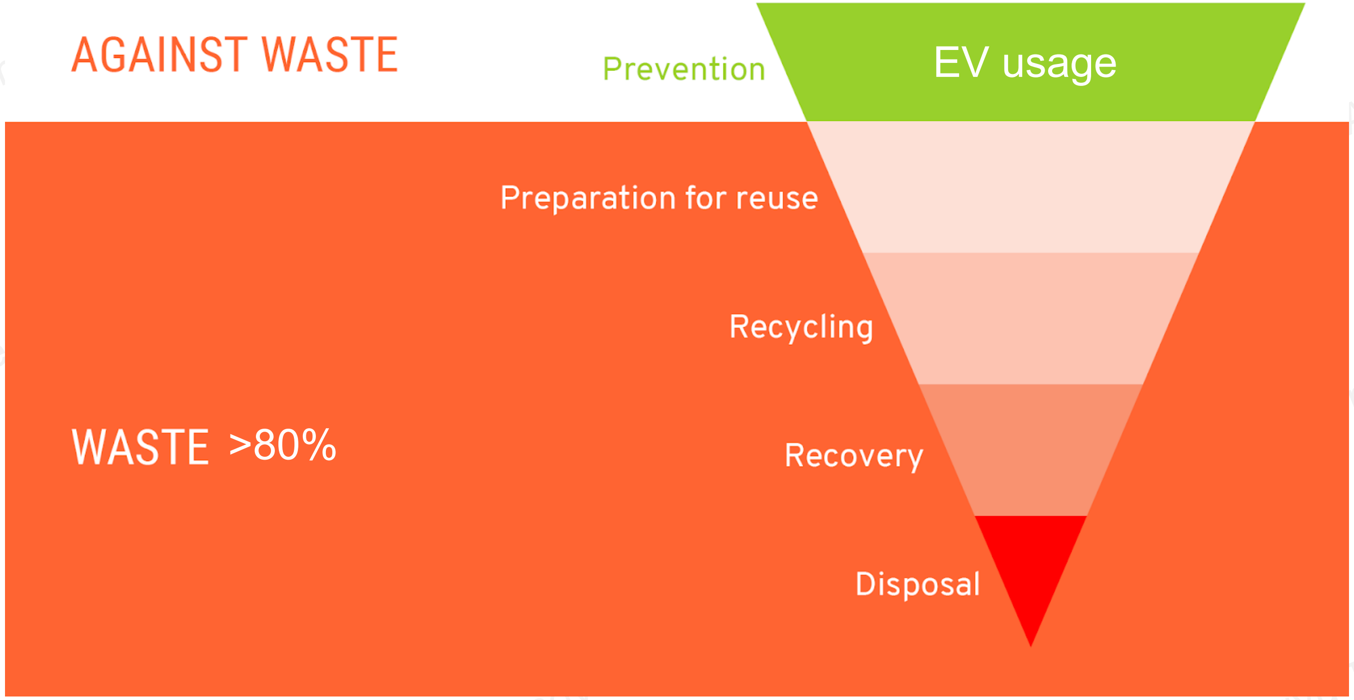


Figure 1.1 The "inverse-pyramid" of electric vehicle energy utilization.

*1 retired 18650 cell = 1m^2 land & 600 liters water wasted
*1 Tesla Model S = 7,104 cells

Figure 1.2 The environmental impact per battery cell.

Chapter 1: Introduction

☐ Background and Motivation

☐ The current situation of unstandardized battery waste treatment (Reuse)

Figure 1.3 The simplified diagram of battery treatment for reuse.

Chapter 1: Introduction

☐ Background and Motivation

☐ The current situation of unstandardized battery waste treatment (Recycling)

Figure 1.4 The simplified diagram of battery treatment for recycling.

Chapter 1: Introduction

□ Background and Motivation

□ The state-of-the-art major policies


中华人民共和国中央人民政府
www.gov.cn
首页 | 简 | 繁 | EN | 登录 | 邮箱 | 无障碍
首页 > 政策 > 国务院政策文件库 > 国务院部门文件　字号：默认 大 超大 | 打印 收藏 留言 |

标　　题：	五部门关于印发《新能源汽车动力蓄电池梯次利用管理办法》的通知	发文机关：	工业和信息化部 科技部 生态环境部 商务部 市场监管总局
发文字号：	工信部联节〔2021〕114号	来　　源：	工业和信息化部网站
主题分类：	工业、交通\其他	公文种类：	通知
成文日期：	2021年08月19日		

工业和信息化部 科技部 生态环境部 商务部 市场监管总局关于印发《新能源汽车动力蓄电池梯次利用管理办法》的通知
工信部联节〔2021〕114号

Efficient & safe reuse

Ministry of Industry and Information Technology
2021 (梯次利用) Reuse


中华人民共和国国家发展和改革委员会
National Development and Reform Commission
热门搜索：油价 产业结构调整指导目录
请输入关键字

首页 > 委属单位话发改

动力电池“退役潮”接踵而至 建立科学规范的回收体系刻不容缓

发布时间：2023/02/28　来源：价格成本调查中心　[打印]

Precise & profitable recycling

National Development and Reform Commission
2023 (回收利用) Recycling


中华人民共和国中央人民政府
www.gov.cn
首页 | 简 | 繁 | EN | 登录 | 邮箱 | 无障碍
首页 > 政策 > 国务院政策文件库 > 国务院文件　字号：默认 大 超大 | 打印 收藏 留言 |

索 引 号：	000014349/2024-00017	主题分类：	城乡建设、环境保护\节能与资源综合利用
发文机关：	国务院办公厅	成文日期：	2024年02月06日
标　　题：	国务院办公厅关于加快构建废弃物循环利用体系的意见		
发文字号：	国办发〔2024〕7号	发布日期：	[illegible]

国务院办公厅关于加快构建
废弃物循环利用体系的意见
国办发〔2024〕7号

Sustainable circulation

General Office of the State Council
2024 (循环利用) Circulated use

Figure 1.5 Major policies of the battery circular economy.

Chapter 1: Introduction

☐ Background and Motivation

☐ **The roles of machine learning**

☐ **Use data to evaluate battery status to inform safe, profitable, sustainable and standardized allocation**

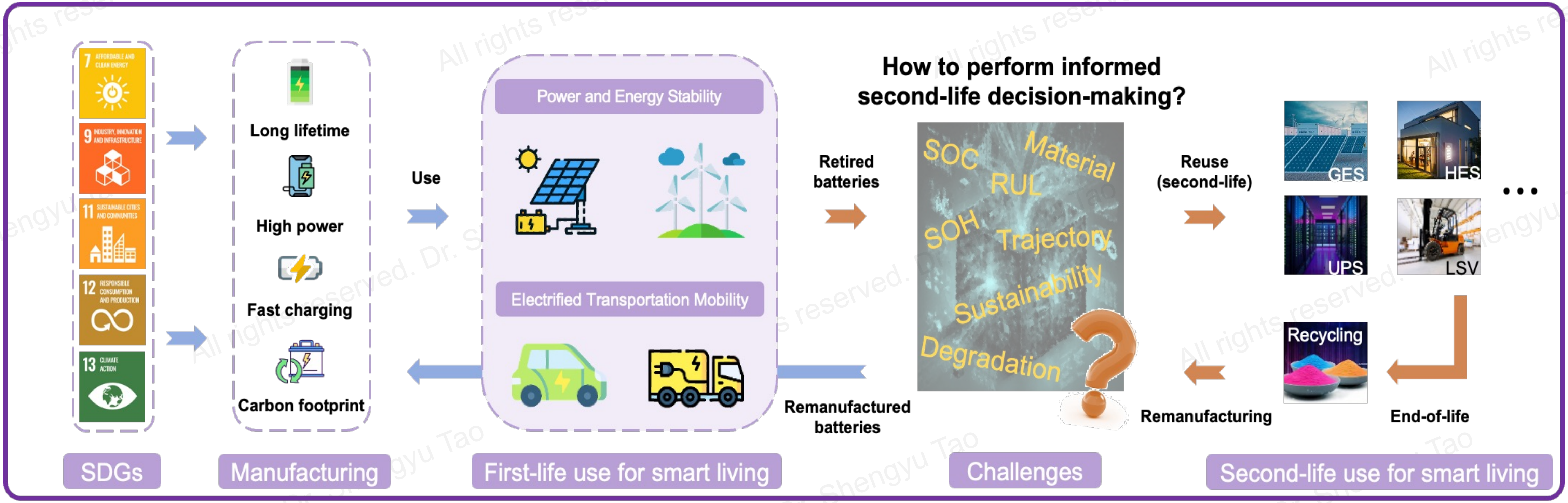
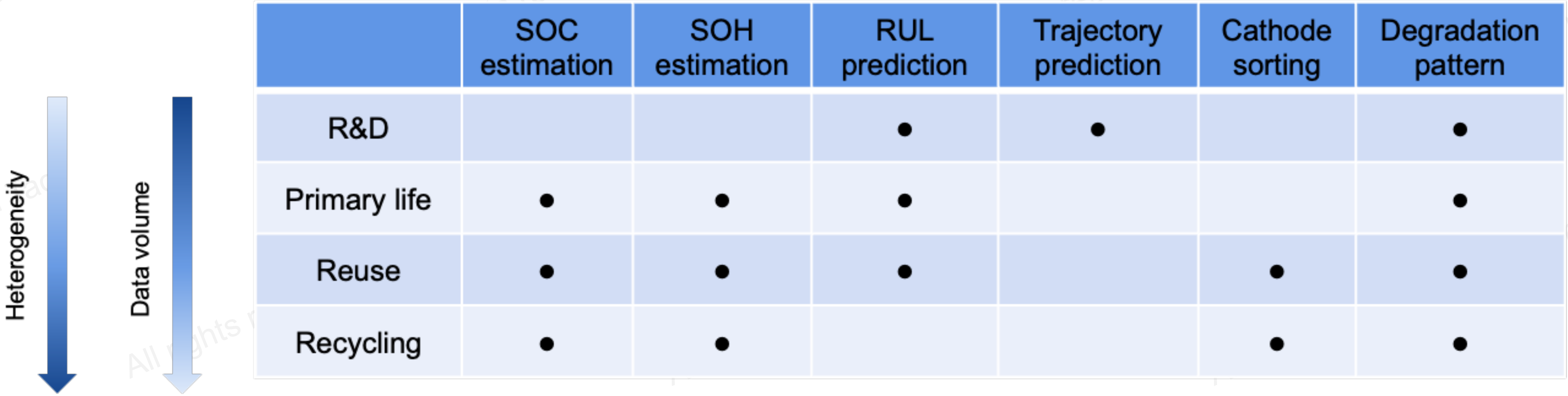


Figure 1.6 The concept of Machine Learning-Assisted Lithium-ion Battery Diagnostics and Prognostics for Sustainable Remanufacturing, Reusing and Recycling.

Chapter 1: Introduction

☐ Background and Motivation

- ☐ The tasks of machine learning
- ☐ The tasks are distinct for different battery service stages



	SOC estimation	SOH estimation	RUL prediction	Trajectory prediction	Cathode sorting	Degradation pattern
R&D			●	●		●
Primary life	●	●	●			●
Reuse	●	●	●		●	●
Recycling	●	●			●	●

● required management task(s)

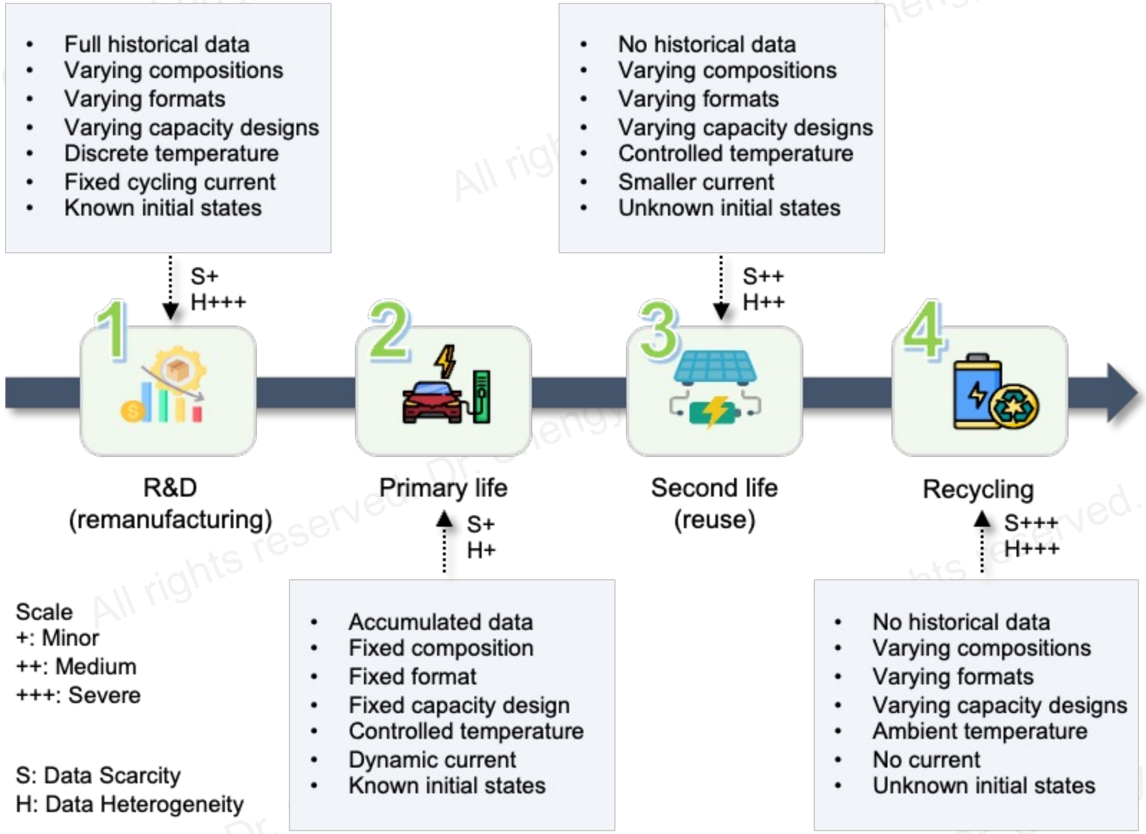
SOC: State of Charge, SOH: State of Health, RUL: Remaining Useful Lifetime

Figure 1.6 The concept of Machine Learning-Assisted Lithium-ion Battery Diagnostics and Prognostics for Sustainable Remanufacturing, Reusing and Recycling.

Chapter 1: Introduction

☐ Literature review

☐ The battery data barriers



- Full historical data
- Varying compositions
- Varying formats
- Varying capacity designs
- Discrete temperature
- Fixed cycling current
- Known initial states

S+
H+++

- No historical data
- Varying compositions
- Varying formats
- Varying capacity designs
- Controlled temperature
- Smaller current
- Unknown initial states

S++
H++

R&D
(remanufacturing)

Primary life

Second life
(reuse)

Recycling

S+
H+

S+++
H+++

Scale
+: Minor
++: Medium
+++: Severe

S: Data Scarcity
H: Data Heterogeneity

- Accumulated data
- Fixed composition
- Fixed format
- Fixed capacity design
- Controlled temperature
- Dynamic current
- Known initial states

- No historical data
- Varying compositions
- Varying formats
- Varying capacity designs
- Ambient temperature
- No current
- Unknown initial states

Figure 1.7 The data characteristics in different battery service stages.

Data scarcity:
Battery status no-longer
monitored post-retirement

Data heterogeneity:
Every battery is different
due to operation varieties

Chapter 1: Introduction

☐ Literature review

☐ The state-of-the-art methods overcoming battery data barriers (data scarcity and heterogeneity)

Data augmentation	Pros	Cons	Ref.
Physical battery testing and data generation	New dataset availability	Time-consuming, costly	[1-5]
Equivalent circuit model	Efficient and affordable	Intensive model priors (biased to fixed patterns)	[6]
Generative AI model (e.g., GAN and VAE)	Data generalization	Overfitting to training distribution	[7-8]
Hybrid model (testing + generative AI)	Balance priors and posteriors	Intensive hyperparameter tuning (another type of model priors)	[9-10]
Model calibration	**Pros**	**Cons**	**Ref.**
Transfer learning	Invariant domain	Theoretical transferability	[11-17]
Continual learning	Continuous model updates	Catastrophic forgetting	[18-20]
Feature engineering (for early prediction)	Efficient	Overfitting and intensive expert knowledge	[1,16,17]

Data augmentation	Pros	Cons	Ref.
Physical battery testing and data generation	New dataset availability	Unexhaustive operation conditions and degradation mechanisms	-
Random charging selection and segment	Simulate real electric vehicle	Not applicable to reuse and recycling	[21-25]
Noise addition	Efficient and affordable	Still adhere to underlying distribution	[26,27]
Model calibration	**Pros**	**Cons**	**Ref.**
Ensemble learning	Mixture of model experts	Intensive hyperparameter tuning and poor interpretability	[28-32]
Transfer learning	Invariant domain	Theoretical transferability	[16,17,33]
Federated learning	Collaborative and private	Communication cost, data missing and poor interpretability	[27, 34]
Large language model	Data generalization	Very early stage and lack of training materials	[35]

Table 1.1 Literature review on state-of-the-art methods overcoming data scarcity and heterogeneity.

Chapter 1: Introduction

☐ Literature review

☐ Research gap

☐ Data scarcity: restricted access, few data…; Data heterogeneity: operation, degradation variabilities…

- Research gaps

- Existing data testing-based generation methods are costly and time-consuming
- Existing prediction methods are vulnerable to varying operation conditions and model parameters

- Solutions:

- Collaborate to maximize existing data source ->mitigate data restrictions
- Generate new condition-constrained data to ensure robustness ->increase data generalizability

- Research gaps

- Existing data generation cannot exhaust case-specific operation conditions
- Existing prediction methods are barely informed of diversified degradation mechanisms

- Solutions:

- Integrate condition information into adaptable diagnostics & prognosis
- Integrate physics (degradation mechanism) into battery diagnostics & prognosis

🟨 Research gaps
⬜ This work

Figure 1.8 The research gap identification.

Chapter 1: Introduction

☐ The high-level work structure

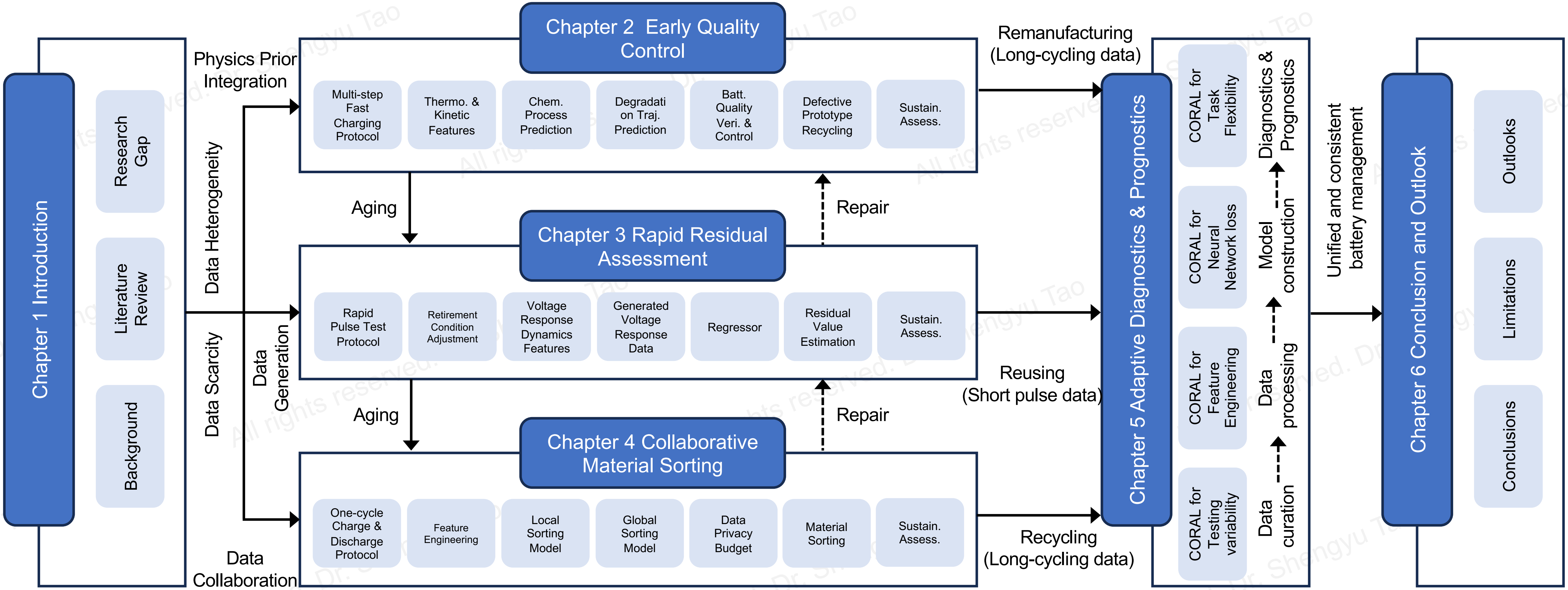


Figure 1.9 The high-level dissertation structure.

Chapter 2: Early Quality Control

Chapter 2: Early Quality Control

☐ Research Question

☐ **How to ensure the quality of remanufactured batteries before put into massive productions?**

- Research gaps

- Existing data testing-based generation methods are costly and time-consuming
- Existing prediction methods are vulnerable to varying operation conditions and model parameters

- Research gaps

- Existing data generation cannot exhaust case-specific operation conditions
- Existing prediction methods are barely informed of diversified degradation mechanisms

- Solutions:

- Collaborate to maximize existing data source ->mitigate data restrictions
- Generate new condition-constrained data to ensure robustness ->increase data generalizability

- Solutions:

- Integrate condition information into adaptable diagnostics & prognosis
- Integrate physics (degradation mechanism) into battery diagnostics & prognosis

🟨 Research gaps

⬜ This work

Chapter 2: Early Quality Control

☐ Overview

☐ **Motivation**

☐ **Traditional quality verification takes > 3 (6) months under accelerated (normal) conditions**

☐ **Accelerated verification are faster but can be hard to translate into normal conditions**

☐ **Contribution**

☐ **Formulates the early quality control as an Arrhenius-transferable time-serials forecasting problem**

☐ **A multi-step fast charging dataset is generated for explainable feature extraction**

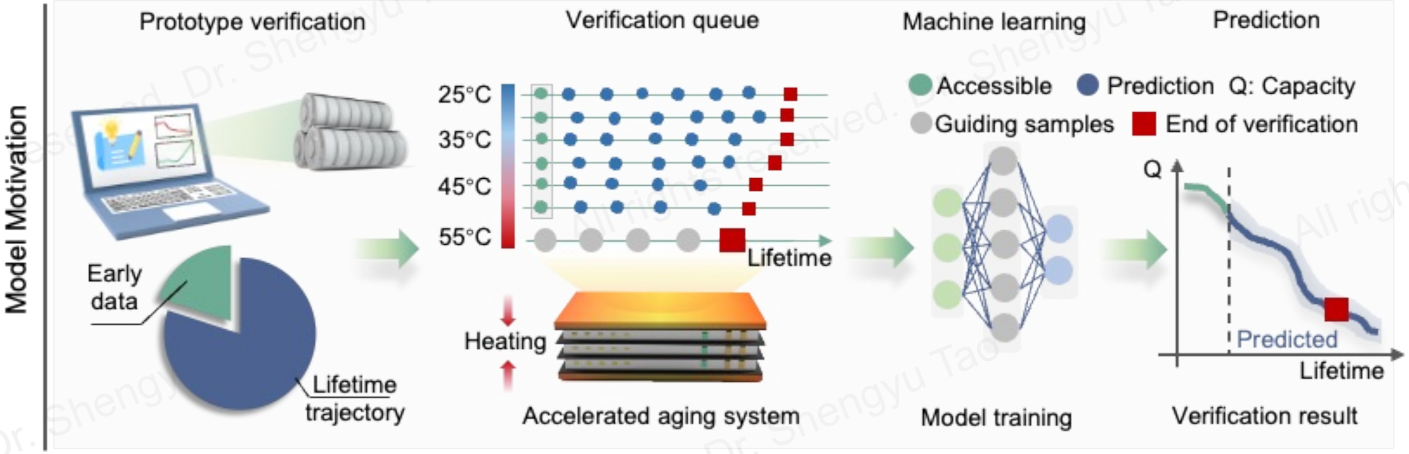


Figure 2.1 The diagram of normal and accelerated quality verification.

Chapter 2: Early Quality Control

☐ Methodology Summary

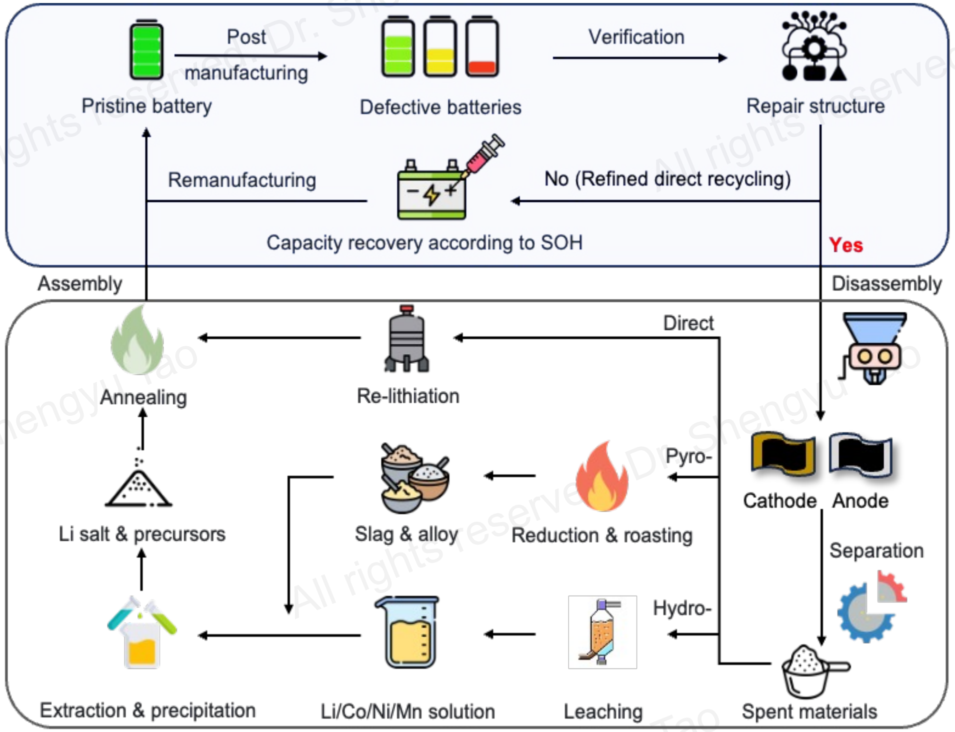
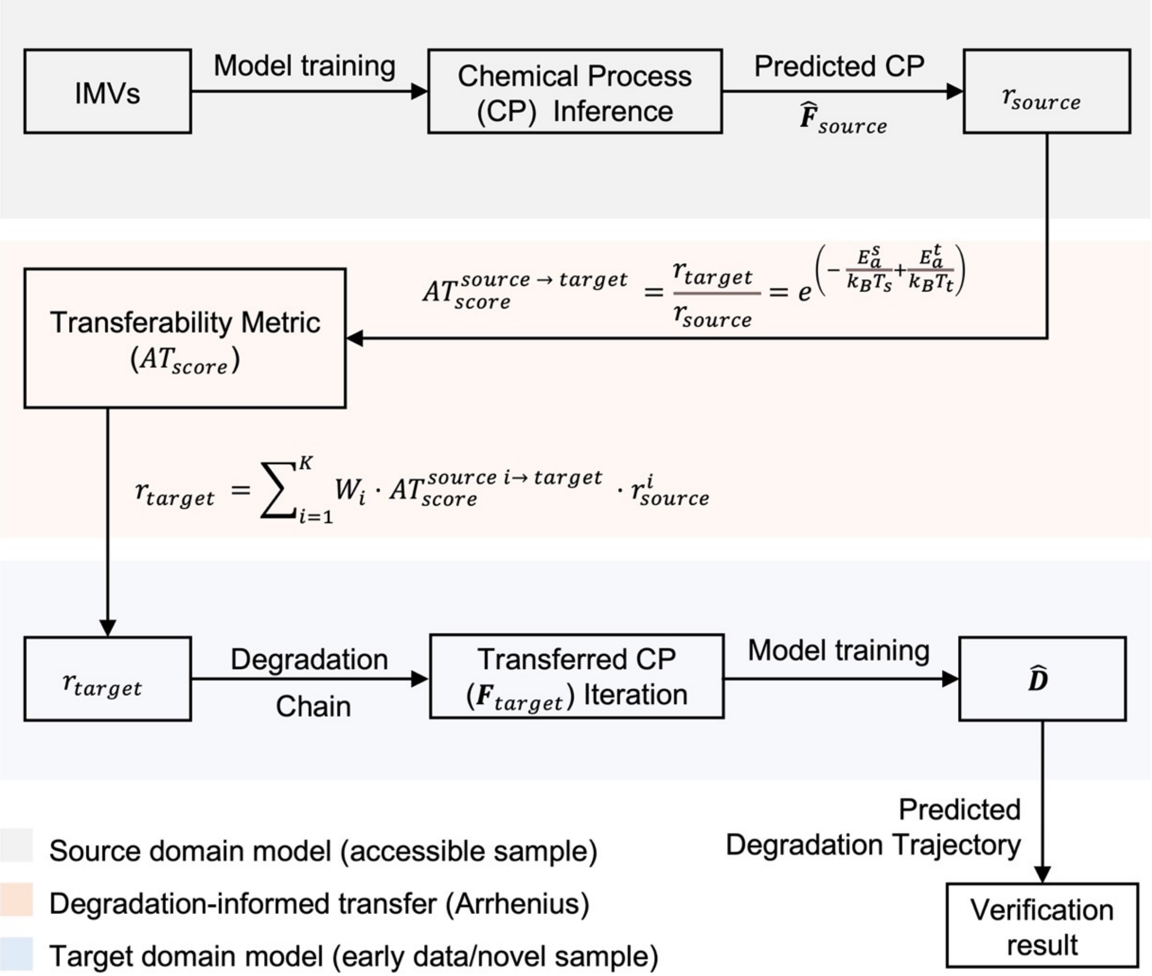


Figure 2.2 The diagram of different recycling methods.



IMVs

Model training

Chemical Process (CP) Inference

Predicted CP $\hat{\boldsymbol{F}}_{source}$

r_{source}

Transferability Metric (AT_{score})

$$AT_{score}^{source \to target} = \frac{r_{target}}{r_{source}} = e^{\left(-\frac{E_a^s}{k_B T_s}+\frac{E_a^t}{k_B T_t}\right)}$$

$$r_{target} = \sum_{i=1}^{K} W_i \cdot AT_{score}^{source\ i \to target} \cdot r_{source}^{i}$$

r_{target}

Degradation Chain

Transferred CP ($\boldsymbol{F}_{target}$) Iteration

Model training

$\hat{\boldsymbol{D}}$

Predicted Degradation Trajectory

Verification result

- ▨ Source domain model (accessible sample)
- ▨ Degradation-informed transfer (Arrhenius)
- ▨ Target domain model (early data/novel sample)

Figure 2.3 The algorithm summary.

Chapter 2: Early Quality Control

☐ Methodology

☐ The Arrhenius equation:

$$r = A\, e^{\left(-\frac{E_a}{k_B T}\right)}$$

where, r is the aging rate of the battery, A is a constant, E_a is the activation energy， k_B is Boltzmann constant and T is the Kelvin temperature.

☐ Arrhenius equation-based transferability metric (AT_{score}):

$$AT_{score}^{source \rightarrow target} = \frac{r_{target}}{r_{source}} = e^{\left(-\frac{E_a^s}{k_B T_s}+\frac{E_a^t}{k_B T_t}\right)} = e^{\left(-\frac{1}{T_s}+\frac{1}{T_t}\right)\alpha}$$

where, T_s and T_t are the Kelvin temperatures of the source domain and the target domain, respectively.

☐ The contribution of source domains: $W_i = \left(\left|AT_{score}^{source\ i \rightarrow target} - 1\right| \cdot \sum_{j=1}^{K} \frac{1}{\left|AT_{score}^{source\ j \rightarrow target} - 1\right|}\right)^{-1}$

where, i is the source domain index.

☐ The aging rate of target domain: $r_{target} = \sum_{i=1}^{K} W_i \cdot AT_{score}^{source\ i \rightarrow target} \cdot r_{source}^{i}$

where, K is the number of source domains.

Chapter 2: Early Quality Control

☐ Methodology

☐ **Chemical Process Prediction Model:** $\hat{\boldsymbol{F}} = f_{\Theta}(U) = \left(f_{\sigma}^{(L)}\left(f_{\theta^{(L)}}^{(L)}\right) \circ \cdots \circ f_{\sigma}^{(1)}\left(f_{\theta^{(1)}}^{(1)}\right)\right)(U)$

where, U (from the charging test) is the broadcasted input voltage matrix $U_{(C\times m)\times 10}$, given a feature matrix $\boldsymbol{F}_{(C\times m)\times N}$.

☐ **Loss of Chemical Process Prediction:** $L_{loss\_ChemicalProcess} = \frac{\sum_{i=1}^{C}(\boldsymbol{F}_i - \hat{\boldsymbol{F}}_i)^2}{C} + \lambda_1 * \sum_{i=1}^{C} |\boldsymbol{F}_i - \hat{\boldsymbol{F}}_i|$

where, U is the broadcasted input voltage matrix $U_{(C\times m)\times 10}$, given a feature matrix $\boldsymbol{F}_{(C\times m)\times N}$, $\boldsymbol{F}_i$ is the ith label of defined chemical processes, $\hat{\boldsymbol{F}}_i$ is the predicted chemical processes feature matrix for the ith cycle, λ_1 is regularization.

☐ **Chemical Process Prediction Model in target domain:** $\boldsymbol{F}_{target}^{i} = \boldsymbol{F}_{target}^{i-1} + \sum_{j=i}^{K} W_j \cdot A_{score}^{source\ j\rightarrow target} \cdot r_{source\,j}^{i-1}$ where, the $\boldsymbol{F}_{target}^{i}$ is the feature value of target domain in the ith cycle, $r_{source\ j}^{i-1}$ is the aging rate of source domain j at the $(i-1)$th cycle .

Chapter 2: Early Quality Control

☐ Methodology

☐ **Degradation Trajectory Prediction Model:** $\hat{D} = f_{\Theta}(\hat{\boldsymbol{F}}) = \left(f_{\sigma}^{(L)}\left(f_{\theta^{(L)}}^{(L)}\right) \circ \cdots \circ f_{\sigma}^{(1)}\left(f_{\theta^{(1)}}^{(1)}\right)\right)(\hat{\boldsymbol{F}})$

where, $\hat{\boldsymbol{F}}$ is the predicted battery chemical process feature matrix.

☐ **Loss of Degradation Trajectory Prediction:** $L_{loss\_DegradationTrajectory} = \frac{\sum_{i=1}^{C}(y_i - \hat{y}_i)^2}{C} + \lambda_2 * \sum_{i=1}^{C}|y_i - \hat{y}_i|$ where, U is the broadcasted input voltage matrix $U_{(C\times m)\times 10}$, given a feature matrix $\boldsymbol{F}_{(C\times m)\times N}$, $\boldsymbol{F}_i$ is the ith label of defined chemical processes, $\hat{\boldsymbol{F}}_i$ is the predicted chemical processes feature matrix for the ith cycle, λ_1 is regularization.

☐ **The SAGE feature importance:** $SAGE = \frac{1}{d}\sum_{S\subseteq D\setminus\{i\}} \binom{d-1}{|S|}^{-1} E[Var(E[Y \mid X_S, X_i] \mid X_S)]$

where, Y is the output of the degradation trajectory prediction model, $X_S \equiv \{X_i \mid i \in S\}$ are subsets of features for different $S \subseteq D$, where D is the set of all features and $D \equiv \{1, ..., d\}$. $\binom{d-1}{|S|}$ equals to combination numbers of features.

Chapter 2: Early Quality Control

Results

- The degradation trajectory prediction achieves low errors while only using 4% data (50 cycles)
- Robust to temperatures, lifetime points, data availabilities, and battery samples

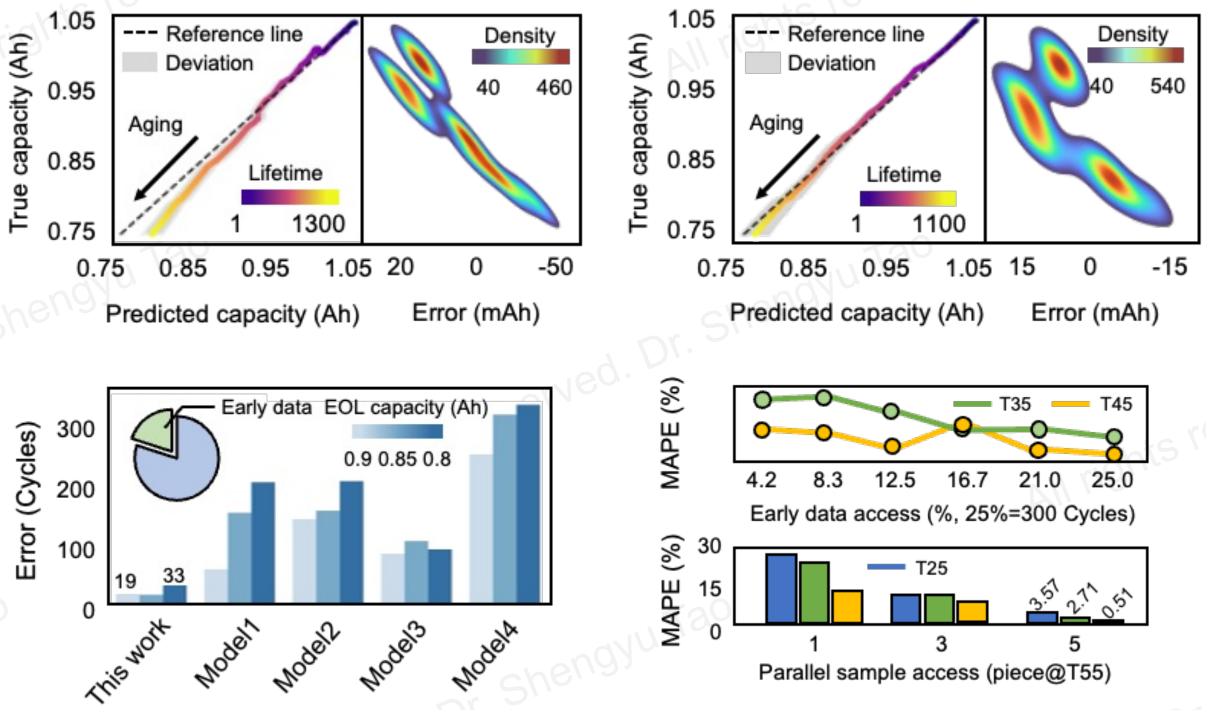


Figure 2.4 The degradation trajectory prediction performances.

	The first 50 cycles are accessible			The first 25 cycles are accessible		
	Verification temperature at 45 degrees					
	MAPE (%)	STD	maxMAPE (%)	MAPE (%)	STD	maxMAPE (%)
Our	0.99	0.36	1.60	1.27	0.44	2.17
Model1	67.75	7.91	81.37	87.95	9.86	104.68
Model2	1.47	0.57	2.72	2.37	0.57	3.44
Model3	7.13	0.64	8.10	7.00	0.57	7.94
Model4	8.78	0.52	9.63	8.83	0.52	9.70
	Verification temperature at 35 degrees					
	MAPE (%)	STD	maxMAPE (%)	MAPE (%)	STD	maxMAPE (%)
Our	2.11	0.73	3.37	2.52	0.80	3.68
Model1	87.99	14.6	107.38	89.78	11.7	107.01
Model2	2.54	0.19	3.57	2.60	0.19	3.80
Model3	11.56	0.59	11.97	11.04	0.79	11.44
Model4	15.66	0.72	16.69	15.89	0.78	17.10
	Verification temperature at 25 degrees					
	MAPE (%)	STD	maxMAPE (%)	MAPE (%)	STD	maxMAPE (%)
Our	2.64	0.82	3.50	3.14	0.85	4.18
Model1	73.66	21.1	108.56	78.94	25.2	138.87
Model2	3.69	0.64	4.79	2.84	0.63	3.87
Model3	9.51	0.91	13.41	11.78	0.82	12.76
Model4	16.83	0.53	17.58	17.22	0.53	18.02

Chapter 2: Early Quality Control

☐ Results

- ☐ **The capacity based features are more important than other manipulated features**
- ☐ **The thermodynamic features are more important than the kinetic features**

  Note: thermodynamic features are those from low current stages, while kinetic ones are from high current stages
- ☐ **The voltage loss (battery degradation) is presented by:**

$$|U_{actual} - U_{theoretical}(*)| = \Delta E(SOC, SOH, T) + \eta(\boldsymbol{I}, SOC, SOH, T)$$

where, U_{actual} is the actual working voltage. $U_{theoretical}$ is the theoretical open-circuit voltage, ΔE is the thermodynamic voltage loss, η is the current-induced kinetic loss.

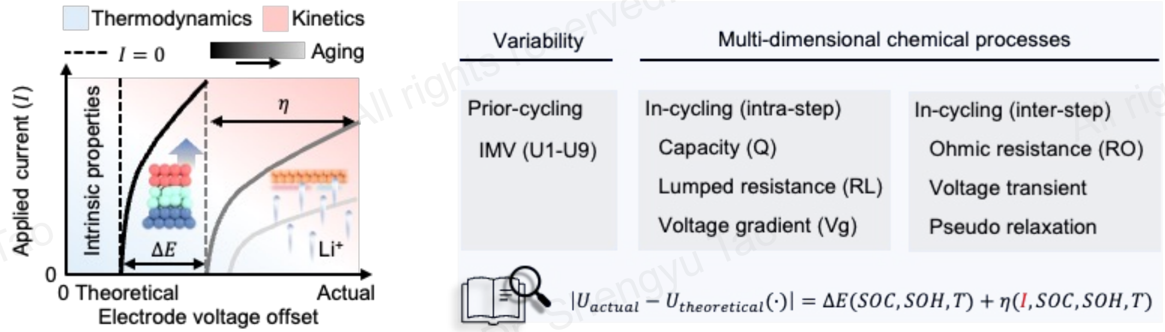


Figure 2.5 The feature engineering methodology.

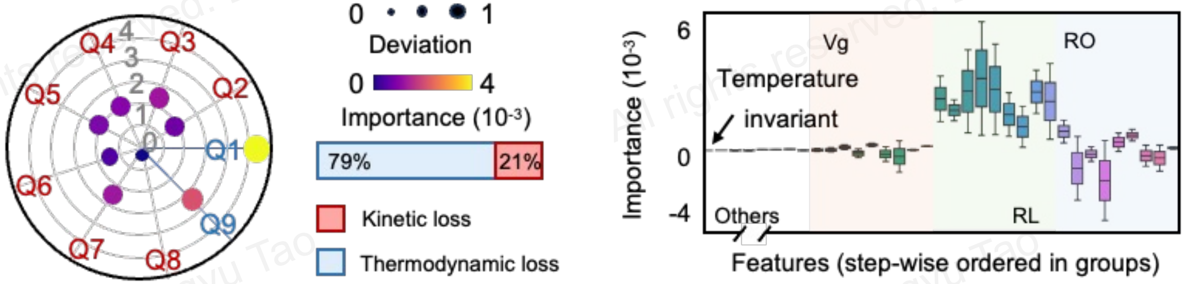


Figure 2.6 The feature importance analysis.

Chapter 2: Early Quality Control

☐ Results

- ☐ **Before the SEI activation, kinetics loss dominates**
- ☐ **At higher temperature, the kinetics loss can lead to higher thermodynamics loss**
- ☐ **The degradation correlation can be used as a non-destructive measure of battery degradation (circle points) to control the battery quality**

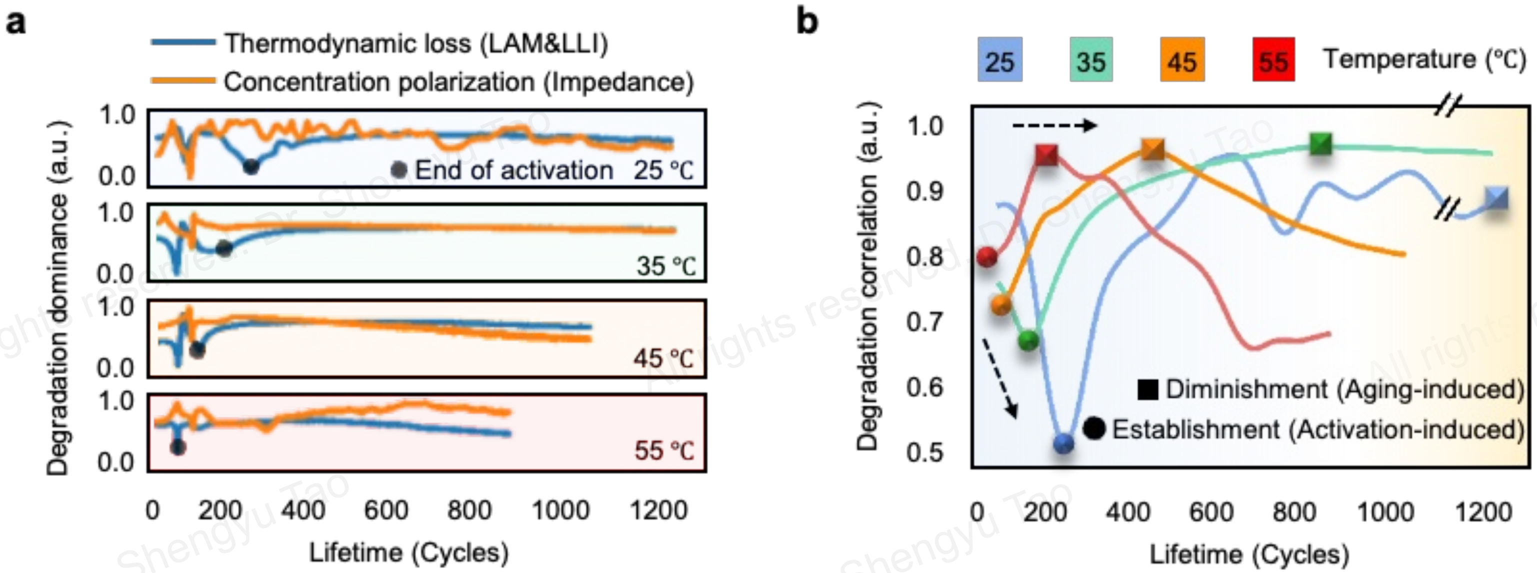


Figure 2.7 The degradation dominance and correlation.

Chapter 2: Early Quality Control

☐ Results

- ☐ **A forward-looking sustainability assessment of “refined-recycling”**
- ☐ **The “refined-recycling” demonstrates a 19.76 billion USD defective material recycling market by 2060 worldwide**
- ☐ **The early quality control is fast, non-destructive and sustainable in treating defective prototype batteries**

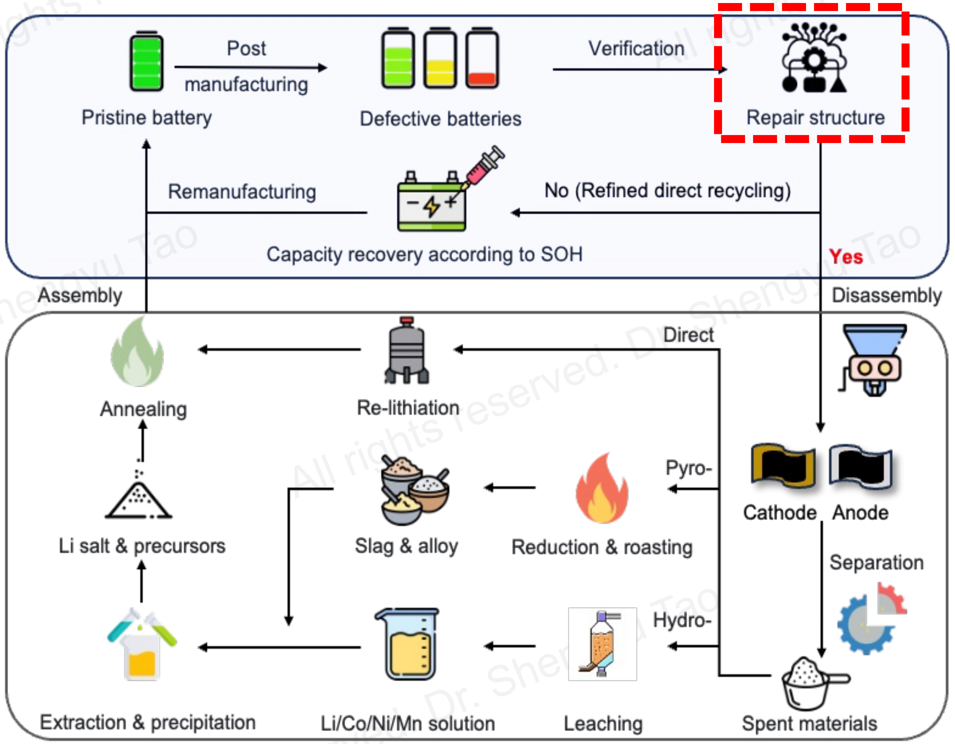


Figure 2.8 The diagram of different recycling methods.

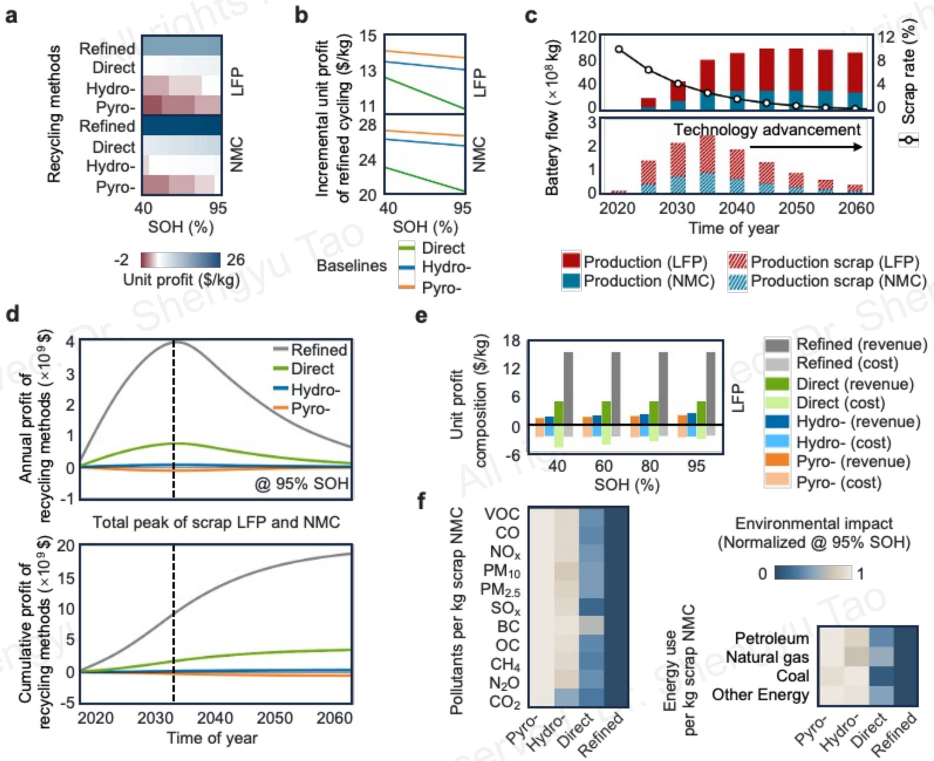


Figure 2.9 The sustainability assessment of different recycling methods.

Chapter 3: Rapid Residual Assessment

Chapter 3: Rapid Residual Evaluation

☐ Research Question

☐ How to ensure the residual evaluation's rapidness and generalizability using minimal data?

- Research gaps

- Existing data testing-based generation methods are costly and time-consuming
- Existing prediction methods are vulnerable to varying operation conditions and model parameters

- Research gaps

- Existing data generation cannot exhaust case-specific operation conditions
- Existing prediction methods are barely informed of diversified degradation mechanisms

- Solutions:

- Collaborate to maximize existing data source ->mitigate data restrictions
- Generate new condition-constrained data to ensure robustness ->increase data generalizability

- Solutions:

- Integrate condition information into adaptable diagnostics & prognosis
- Integrate physics (degradation mechanism) into battery diagnostics & prognosis

🟨 Research gaps

⬜ This work

Chapter 3: Rapid Residual Assessment

☐ Overview

☐ **Motivation**

☐ **The residual value evaluation requires extensive data measurement**

☐ **The data measurement are sensitive to different state of charge conditions and material types**

☐ **Contribution**

☐ **A real-world pulse test based residual value assessment dataset is generated**

☐ **A condition-dependable generative model is built for extended pulse measurement data generation**

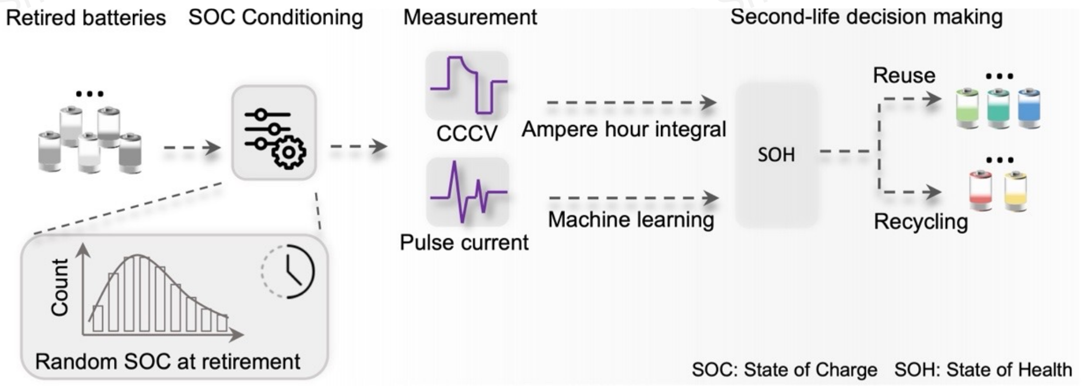


Figure 3.1 The diagram of rapid residual assessment.

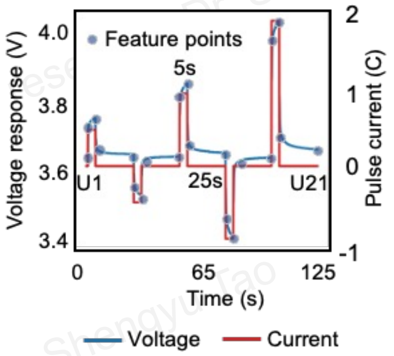


Figure 3.2 Pulse test.

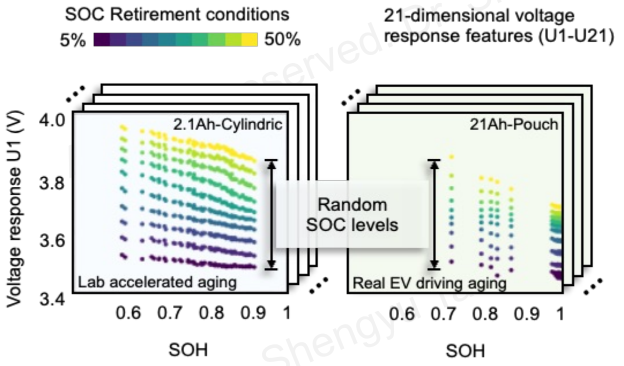


Figure 3.3 Condition dependency of pulse test.

Chapter 3: Rapid Residual Assessment

☐ Methodology Summary

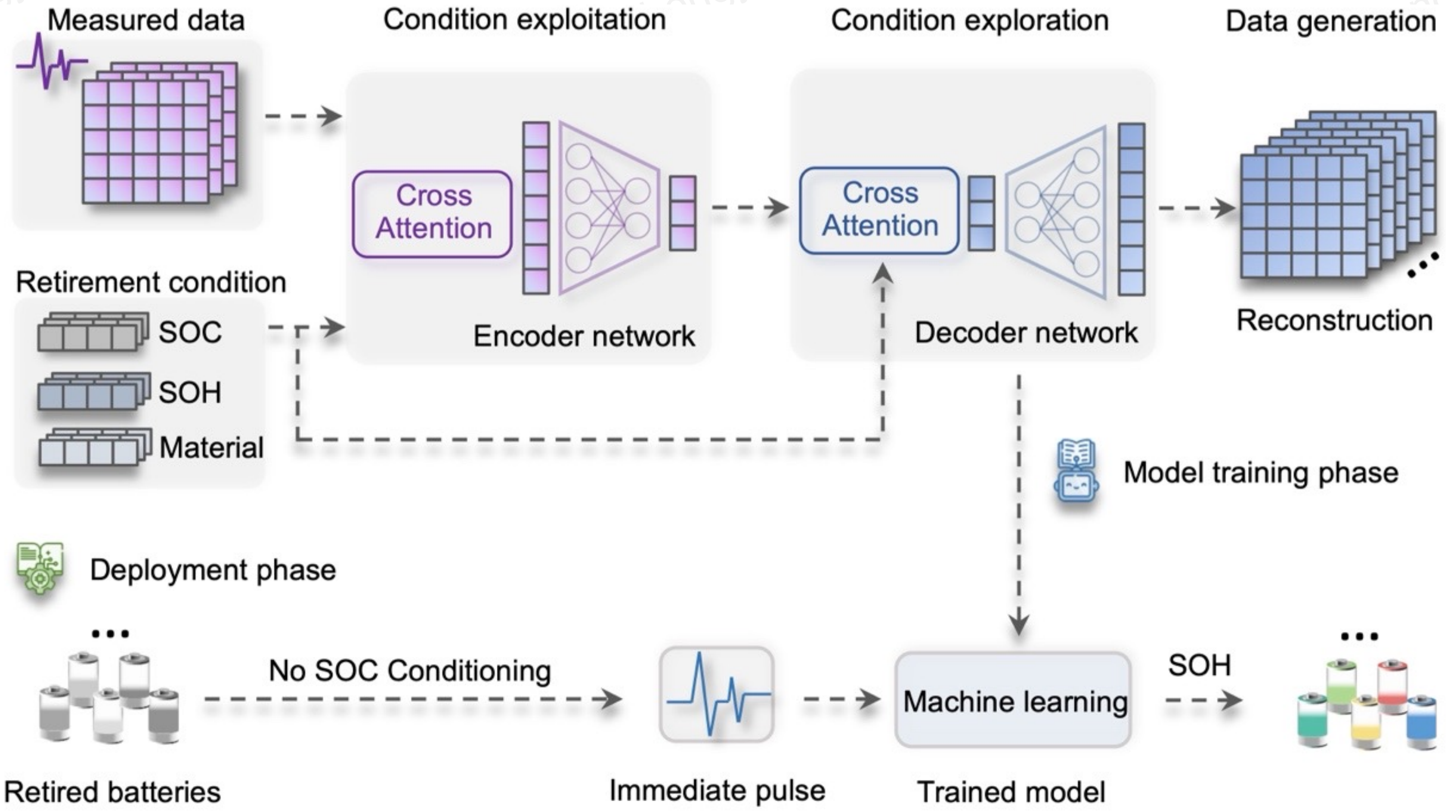


Figure 3.3 The algorithm summary.

Chapter 3: Rapid Residual Assessment

☐ Methodology

☐ **The attention mechanism:** $\text{Attention}\,(Q,K,V)=\text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$

where, Q,K,V represent the query, key, and value sequence, respectively. d_k is the scaling factor, typically the dimension value of the key K. T is the transpose operator. The softmax function normalizes the input vector so that the sum of the probabilities is 1.

☐ **The conditional embedding for the encoder:** $\mathbf{C}=\text{ReLU}\left(\mathbf{cond}\cdot\mathbf{W}_\mathbf{c}^\mathbf{T}+\mathbf{b}_\mathbf{c}\right)$

where, $\mathbf{cond}=[SOC,SOH]\in\mathbb{R}^{N\times2}$ is the condition input, $\mathbf{W}_\mathbf{c}\in\mathbb{R}^{64\times2}$, $\mathbf{b}_\mathbf{c}\in\mathbb{R}^{\mathrm{N}\times64}$ are the condition embedding neural network weighting matrix and bias matrix.

☐ **The feature embedding for the encoder:** $\mathbf{H}=\text{ReLU}\left(\mathbf{x}\cdot\mathbf{W}_\mathbf{h}^\mathbf{T}+\mathbf{b}_\mathbf{h}\right)$

where, $\mathbf{x}\in\mathbb{R}^{N\times21}$ is the feature matrix input, $\mathbf{W}_\mathbf{h}\in\mathbb{R}^{64\times21}$, $\mathbf{b}_\mathbf{h}\in\mathbb{R}^{\mathrm{N}\times64}$ are the main input embedding neural network weighting matrix and bias matrix.

Chapter 3: Rapid Residual Assessment

☐ Methodology

☐ **The latent embedding after the cross-attention of encoder:** $\mathbf{AttenEncoder} = \text{Attention}\ (\mathbf{H}, \mathbf{C}, \mathbf{C})$

where, $\mathbf{H} \in \mathbb{R}^{N\times 64}$ and $\mathbf{C} \in \mathbb{R}^{N\times 64}$ are embeddings of pulse voltage response data and retirement condition data, respectively.

☐ **The parameter for the latent space of the variational auto-encoder:**

mean of Gaussian distribution: $\mathbf{z_{mean}} = \mathbf{AttenEncoder} \cdot \mathbf{W}_{\mathbf{z_{mean}}}^{\mathbf{T}} + \mathbf{b}_{\mathbf{z_{mean}}}$

variance of Gaussian distribution: $\mathbf{z_{log\_var}} = \mathbf{AttenEncoder} \cdot \mathbf{W}_{\mathbf{z_{log\_var}}}^{\mathbf{T}} + \mathbf{b}_{\mathbf{z_{log\_var}}}$

☐ **Sampling in the latent space:** $\mathbf{z} = \mathbf{z_{mean}} + e^{\frac{1}{2}\cdot \mathbf{z_{log\_var}}} \cdot \boldsymbol{\epsilon}$

☐ **The feature embedding from the decoder:** $\mathbf{H}' = \text{ReLU}\left(\mathbf{z} \cdot \mathbf{W}_{\mathbf{d}}^{\mathbf{T}} + \mathbf{b_d}\right)$

☐ **The latent embedding after the cross-attention of decoder: :** $\mathbf{AttenDecodeder} = \text{Attention}\ (\mathbf{H}', \mathbf{C}', \mathbf{C}')$

☐ **The reconstructed feature matrix output:** $\hat{\mathbf{x}} = \sigma\left(\mathbf{AttenDecodeder} \cdot \mathbf{W}_{\mathbf{o}}^{\mathbf{T}} + \mathbf{b_o}\right)$

Chapter 3: Rapid Residual Assessment

☐ Methodology

- **The reconstruction loss of variational autoencoder:** $\mathrm{Loss}_{\mathrm{MSE}} = \frac{1}{N}\sum_{i=1}^{N}(\mathbf{x}_{\mathbf{i}} - \hat{\mathbf{x}}_{\mathbf{i}})^2$

- **The KL-divergence loss of variational autoencoder:** $\mathrm{Loss}_{\mathrm{KL}} = -\frac{1}{2}\sum_{i=1}^{N}\left(1 + \mathbf{z}_{\mathbf{log\_var}_{\mathbf{i}}} - \mathbf{z}_{\mathbf{mean}_{\mathbf{i}}}^2 - \mathrm{e}^{\mathbf{z}_{\mathbf{log\_var}_{\mathbf{i}}}}\right)$

- **The total loss of variational autoencoder:** $\mathrm{Loss} = \omega_{\mathrm{xent}} \cdot \mathrm{Loss}_{\mathrm{MSE}} + \omega_{\mathrm{KL}} \cdot \mathrm{Loss}_{\mathrm{KL}}$

  where, ω_{xent} and ω_{KL} are set to 0.5 to achieve a balance between the generation accuracy and the diversity, respectively. N is the sample size.

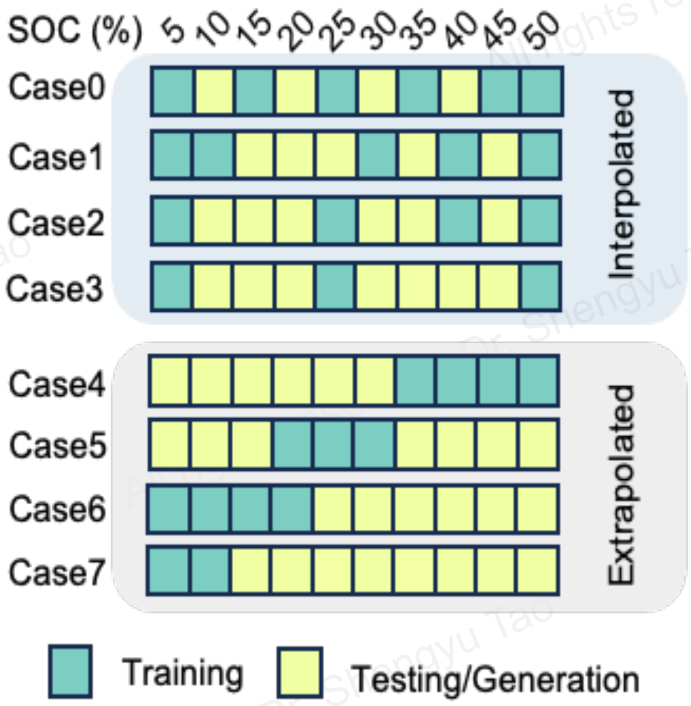
- **The random forest regressor:** $\bar{\mathbf{y}} = \bar{h}(\mathbf{X}) = \frac{1}{K}\sum_{m=1}^{M} h(\mathbf{X}; \boldsymbol{\vartheta}_{\mathrm{m}}, \boldsymbol{\theta}_{\mathrm{m}})$

  where, $\mathbf{X}$ is the generated feature matrix, $\bar{\mathbf{y}}$ is the predicted SOH value vector. M is the tree number in the random forest. $\boldsymbol{\vartheta}_{\mathrm{m}}$ and $\boldsymbol{\theta}_{\mathrm{m}}$ are the hyperparameters.

Chapter 3: Rapid Residual Assessment

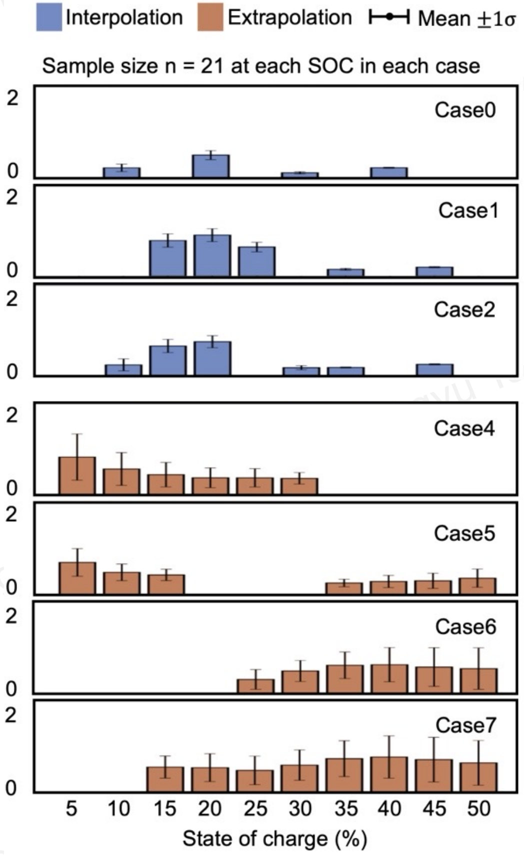
☐ Results

- ☐ **The data generation performance are below 2% MAPE.**
- ☐ **The generation saves time and energy!**



interpolate

extrapolate



Data is cheaper and greener!

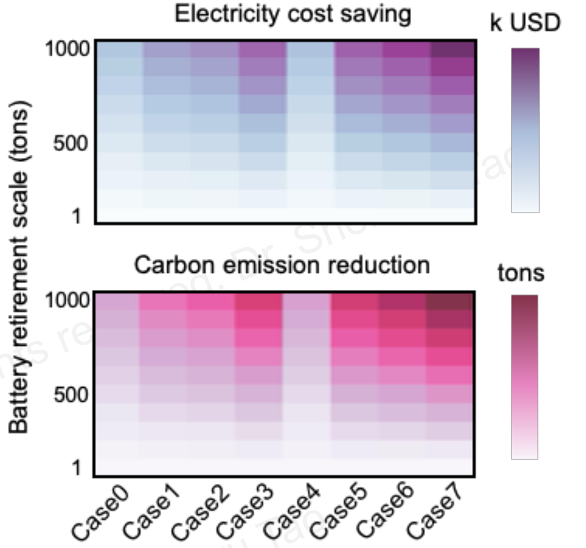


Figure 3.4 The data training and testing split.

Figure 3.5 Data generation results of all cases.

Figure 3.6 Cost and carbon savings.

Chapter 3: Rapid Residual Assessment

☐ Results

- ☐ **The residual assessment is rapid and accurate using generated data, which is necessary.**
- ☐ **The data generation can be generalized to other battery chemistries and formats.**

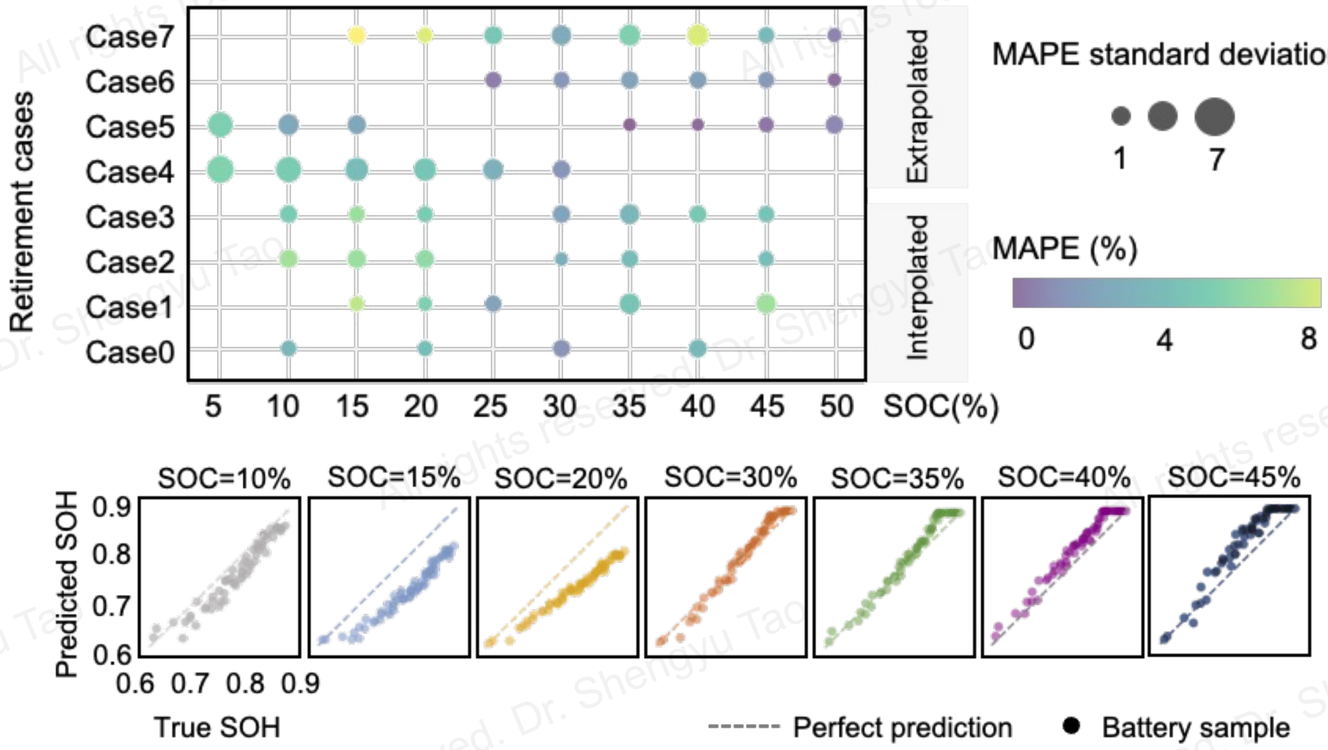


Figure 3.7 The residual assessment performance across all cases.

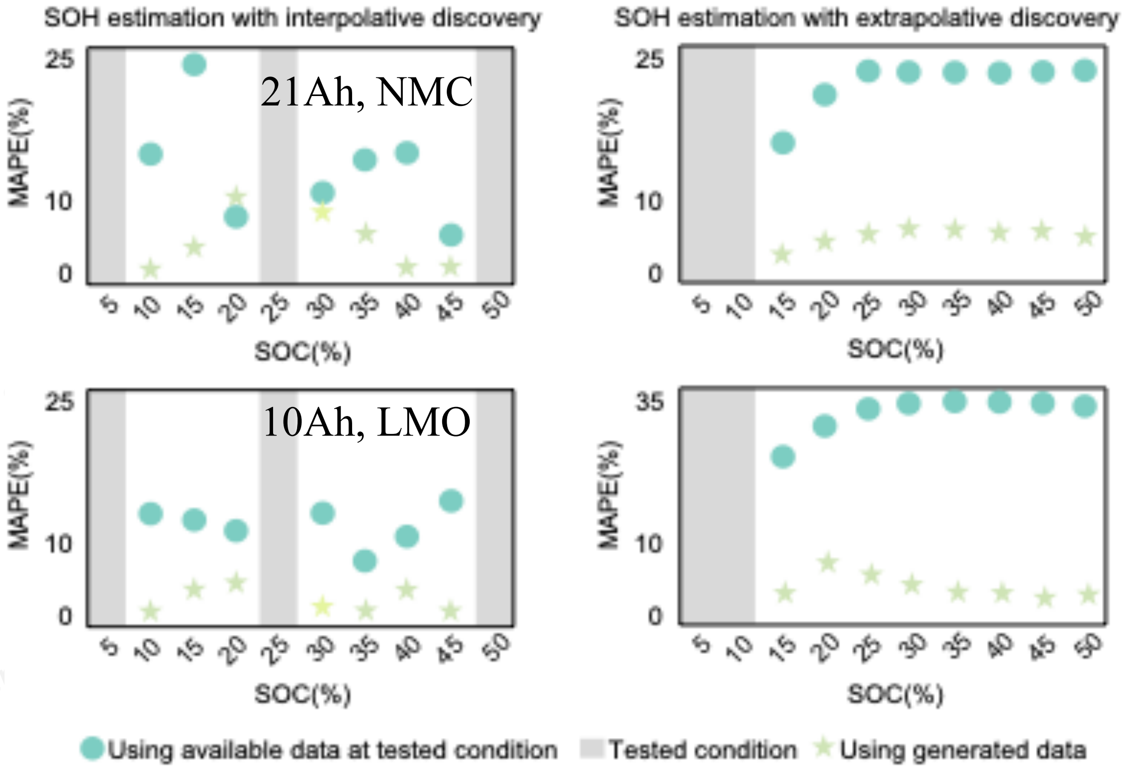
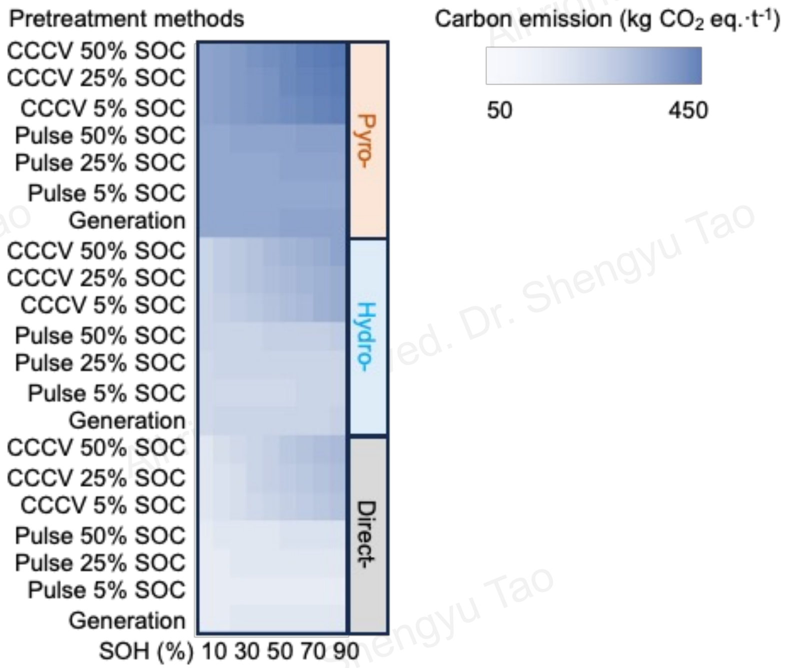


Figure 3.8 Comparison between original and generated data.

Chapter 3: Rapid Residual Assessment

☐ Results

- ☐ **The data generation involves least carbon emission among all typical recycling methods.**
- ☐ **The data generation reduces the pretreatment electricity cost over 50%.**



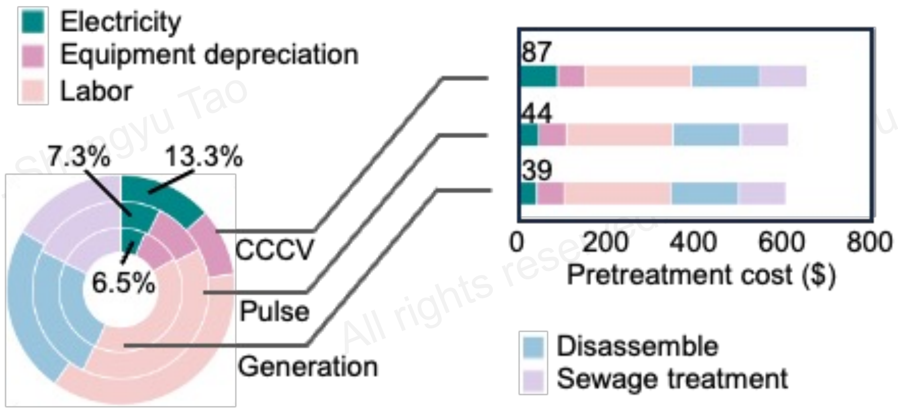


Figure 3.9 The carbon comparison of different residual assessment methods.

Figure 3.10 The cost comparison of different residual assessment methods.

Chapter 4: Collaborative Material Sorting

Chapter 4: Collaborative Material Sorting

☐ Research Question

☐ **How to identify the material status of retired batteries using multi-entity data (with data privacy)?**

- Research gaps

- Existing data testing-based generation methods are costly and time-consuming
- Existing prediction methods are vulnerable to varying operation conditions and model parameters

- Research gaps

- Existing data generation cannot exhaust case-specific operation conditions
- Existing prediction methods are barely informed of diversified degradation mechanisms

- Solutions:

- Collaborate to maximize existing data source ->mitigate data restrictions
- Generate new condition-constrained data to ensure robustness ->increase data generalizability

- Solutions:

- Integrate condition information into adaptable diagnostics & prognosis
- Integrate physics (degradation mechanism) into battery diagnostics & prognosis

🟨 Research gaps

⬜ This work

Chapter 4: Collaborative Material Sorting

☐ Overview

☐ Motivation

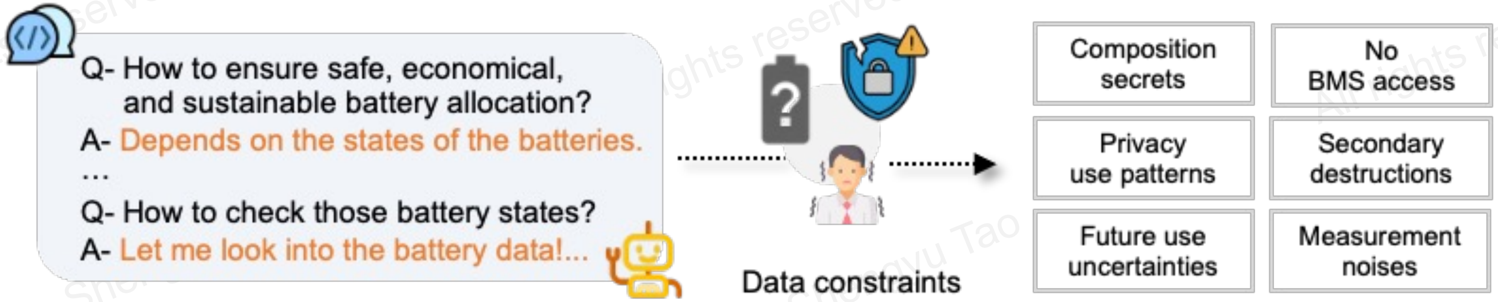
☐ **The battery material data can involve commercial secrets and privacy use patterns.**

☐ **The battery materials are often mixed when collected from different vendors.**

☐ Contribution

☐ **A collaborative and decentralized material sorting algorithm is proposed.**

☐ **The algorithm can protect data privacy by using the data locally while not sharing them globally.**



Q- How to ensure safe, economical, and sustainable battery allocation?

A- Depends on the states of the batteries.

…

Q- How to check those battery states?

A- Let me look into the battery data!...

Figure 4.1 The data privacy constraints in material property identification.

Chapter 4: Collaborative Material Sorting

☐ Methodology Summary

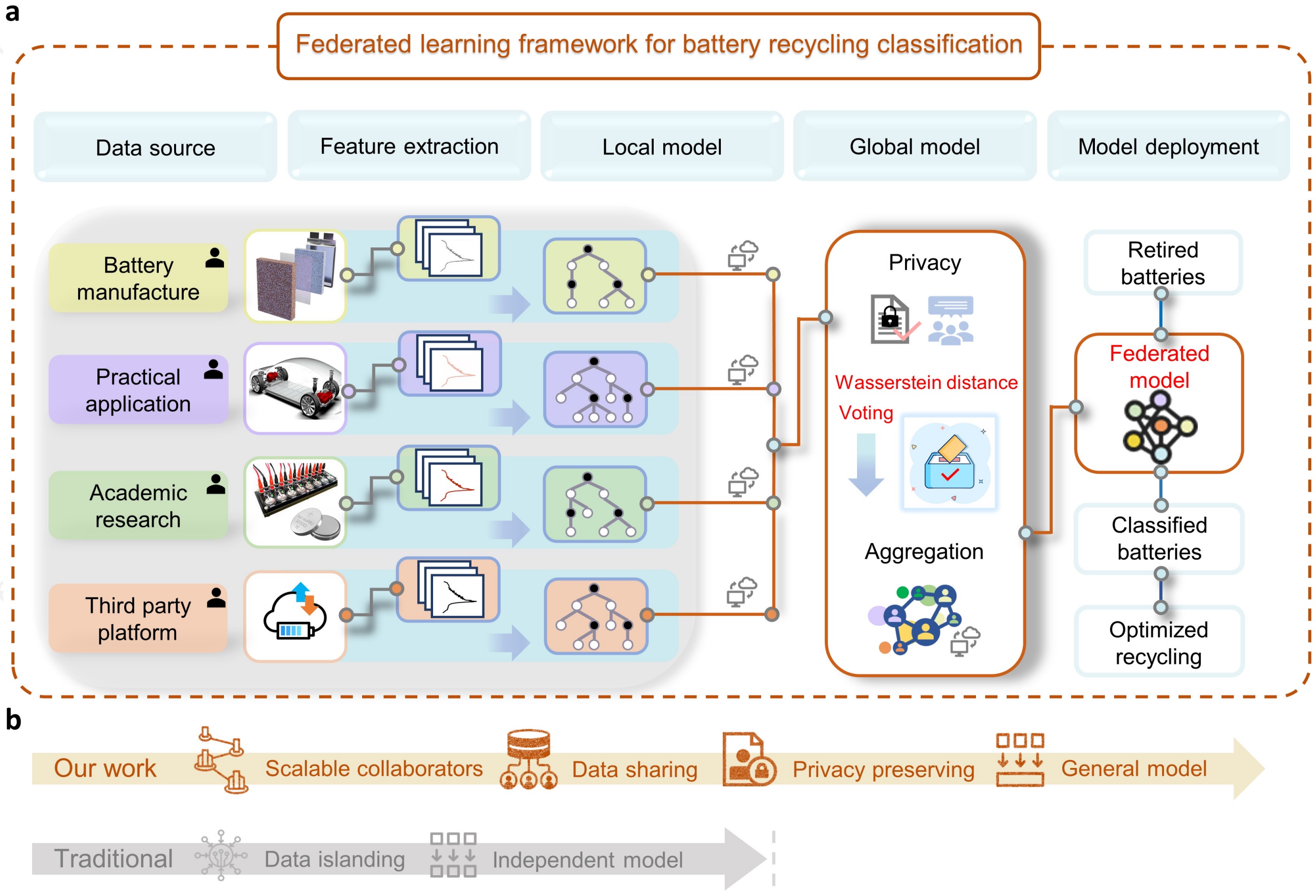


Figure 4.2 The algorithm summary.

Chapter 4: Collaborative Material Sorting

☐ Methodology

☐ **The client (one data collaborator):** $\text{Client}_k \triangleq (\text{lb}_k, \text{ub}_k, \text{NC}_k)$

where, $k = \{1,2,\cdots,10\}$, where lb_k and ub_k are the minimum and maximum number of battery observations in Client_k. NC_k stands for the minimum number of classes in each Client_k.

☐ **The client decision (classification) model:** $g(\mathbf{X}) = \underset{y \in \tilde{C}}{\text{argmax}} \sum_{j=1}^{J} I\left(y = h_j(\mathbf{X};\ \Theta_j)\right)$

where, I is the indicator function. $I\left(y = h_j(\mathbf{X};\ \Theta_j)\right) = 1$ if $y = h_j(\mathbf{X};\ \Theta_j)$ and 0 otherwise.

☐ **The Wasserstein distance measure:** $W_q(\cdot,\cdot) = \left(\inf_{\gamma \in \text{MP}} \int_{\Omega_1 \times \Omega_2} |x_1 - x_2|^q d\gamma(x_1, x_2)\right)^{\frac{1}{q}}$

where, where γ is a transport operator, referring to the transport of arbitrary attributes pairs, i.e., (x_1, x_2), from the global feature space Ω_1 to the client feature space Ω_2. MP stands for a measurably preserved transport.

Chapter 4: Collaborative Material Sorting

☐ Methodology

☐ **The Wasserstein Distance Voting term:** $\omega_k = \alpha^{-\lambda(1-\mathcal{M}_k(W_q))}$

where, $\alpha > 0$ and $\lambda > 0$ are voting hyperparameters. $\mathcal{M}_k$ is the average operator on the pairwise Wasserstein distance between feature spaces of Client_k and the global feature space, i.e., the recycler.

☐ **The aggregated global model:** $G(x) = \underset{y \in \tilde{C}}{\operatorname{argmax}} \sum_{k=1}^{K} \omega_k I\big(y = g_k(x)\big)$

where, K is the number of clients. I is the indicator function. $I\big(y = h_j(\mathbf{X};\ \Theta_j)\big) = 1$ if $y = h_j(\mathbf{X};\ \Theta_j)$ and 0 otherwise.

☐ **The permutation feature importance:** $\text{Imp}_n^m = \frac{1}{J_n}\sum_{j\in\mathcal{J}_n} I\big(y_n \neq \hat{C}^*_{n,j}\big) - \frac{1}{J_n}\sum_{j\in\mathcal{J}_n} I\big(y_n \neq \hat{C}_{n,j}\big)$

where, $\mathcal{J}_n$ is the cardinality of the n out-of-bag observations, J_n is the number of trees in the random forest considering n out-of-bag observations. The feature importance of Imp_n^m is averaged over all observations as a global importance.

☐ **The privacy budget:** $\text{PB} = 100\% \times \max\big(\text{NSR} \big| A(\text{NSR}) \geq \underline{A}\big)$

where, $\underline{A}$ denotes the lower bound of acceptable accuracy of the model, NSR is the noise-to-signal ratio.

Chapter 4: Collaborative Material Sorting

☐ Results

- ☐ **The input for material sorting only requires one cycle (random) end-of-life data.**
- ☐ **The feature extraction process is standardized for different material types .**

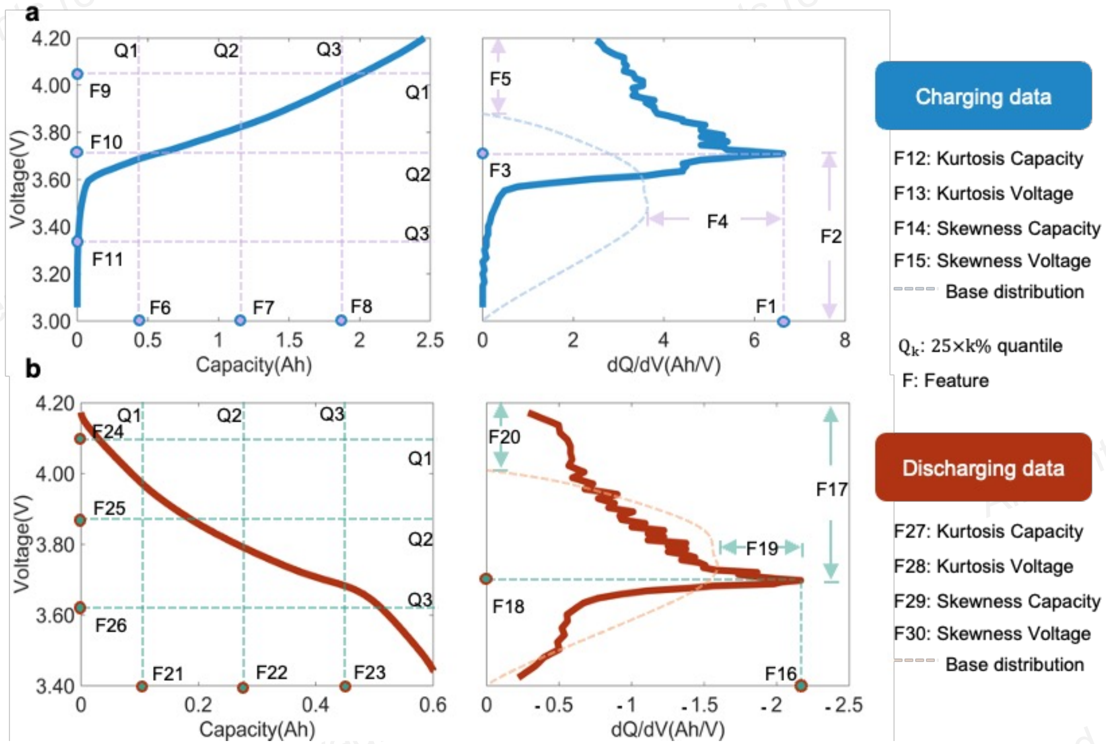


Figure 4.3 The feature engineering.

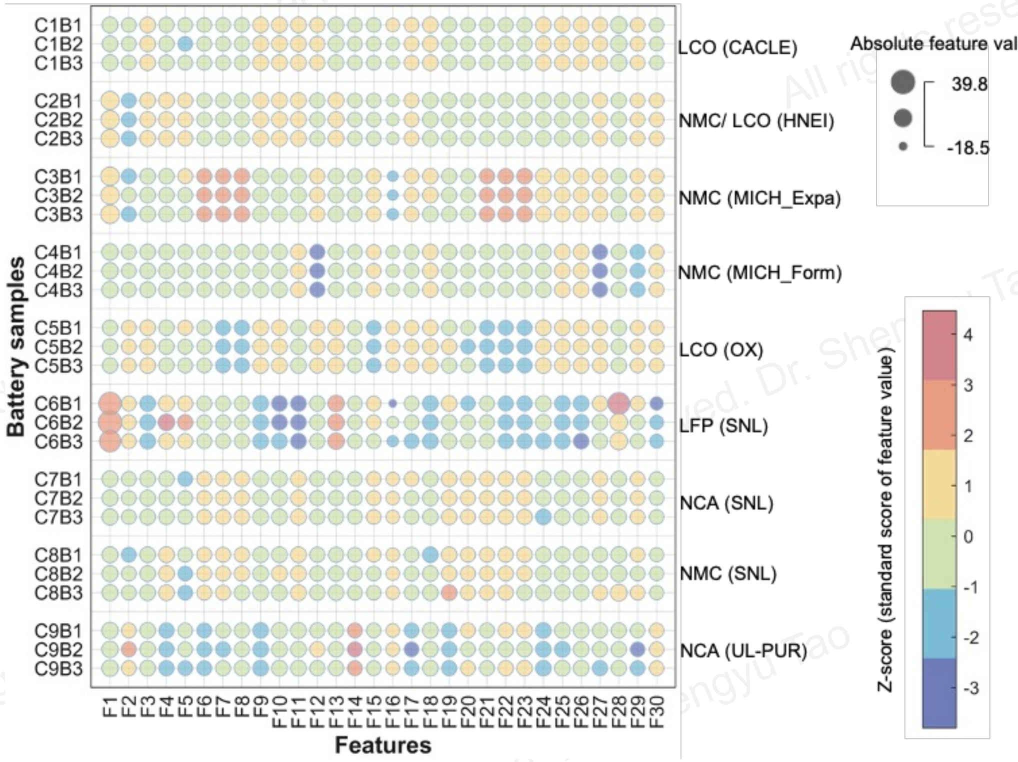


Figure 4.4 The extracted features.

Chapter 4: Collaborative Material Sorting

☐ Results

- ☐ **The proposed WDV aggregation strategy achieve 99% sorting accuracy with homogeneous data access.**
- ☐ **The proposed WDV aggregation strategy is robust to strong measurement noise (up to 10%), thus having the best privacy budget (accuracy > 90% with 10% random noise injection).**

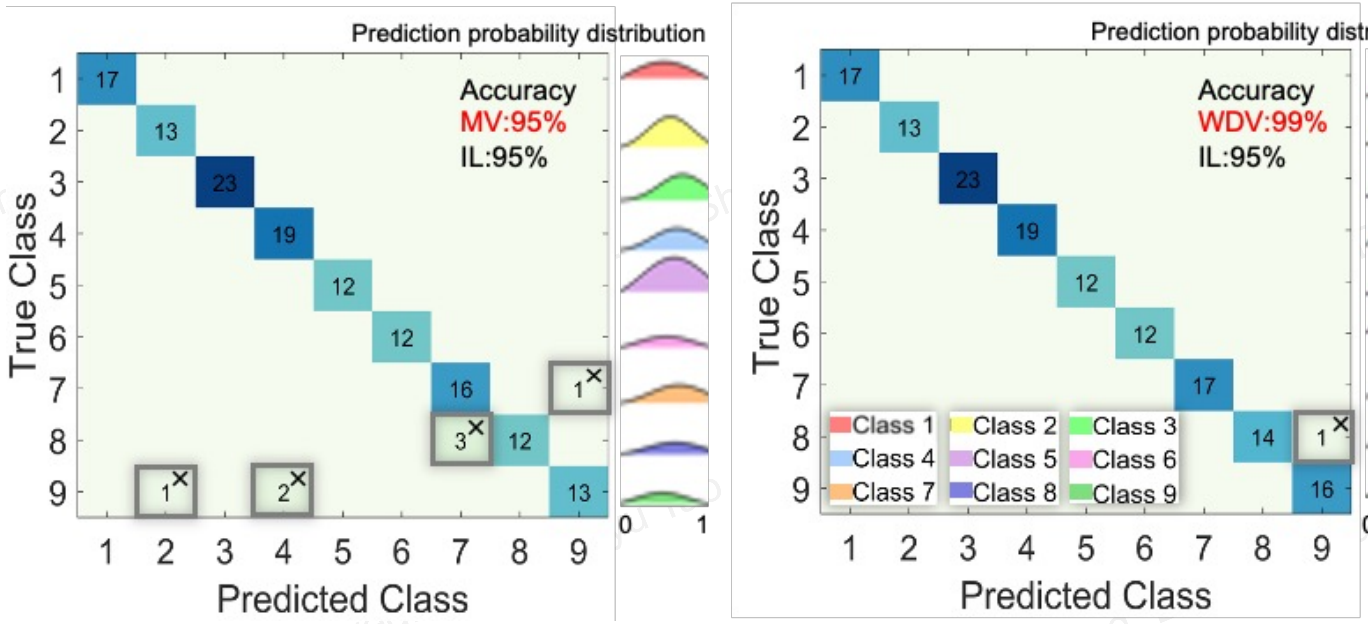


Figure 4.5 Material sorting results with homogenous data access.

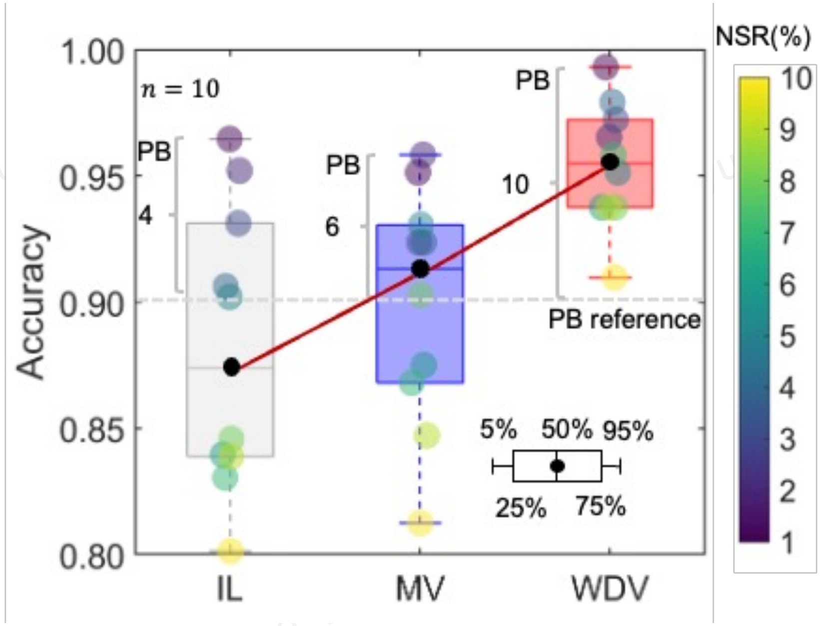


Figure 4.6 Material sorting results with data privacy budgets.

Chapter 4: Collaborative Material Sorting

Results

- **The proposed WDV aggregation strategy outperforms among all heterogeneity indexes.**
- **The proposed WDV aggregation strategy increased 26% of the material sorting accuracy compared to the non-collaborative sorting method.**

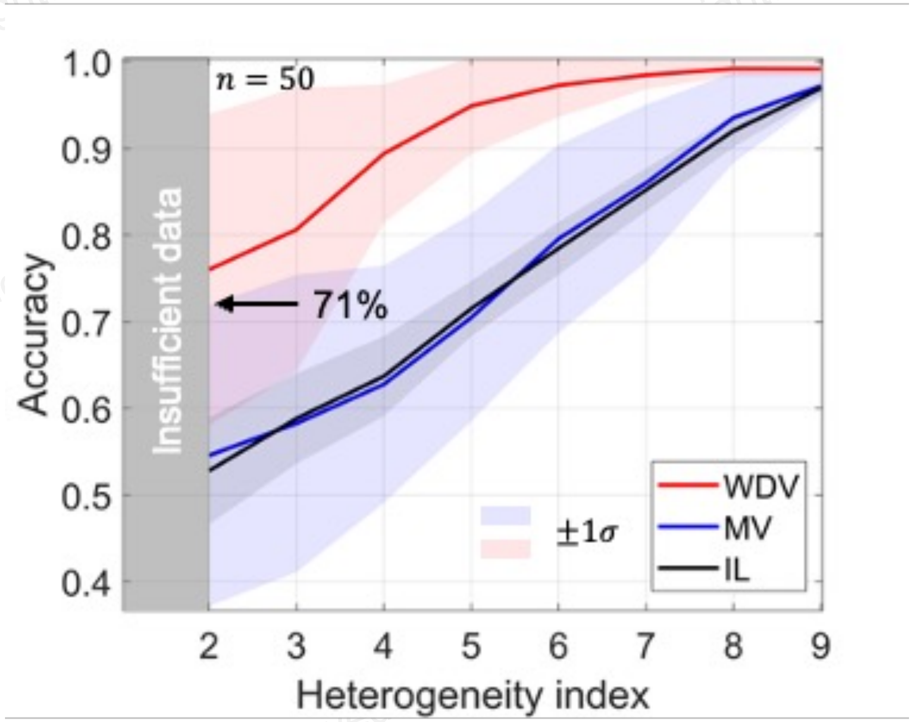


Figure 4.7 Material sorting results with heterogenous data access.

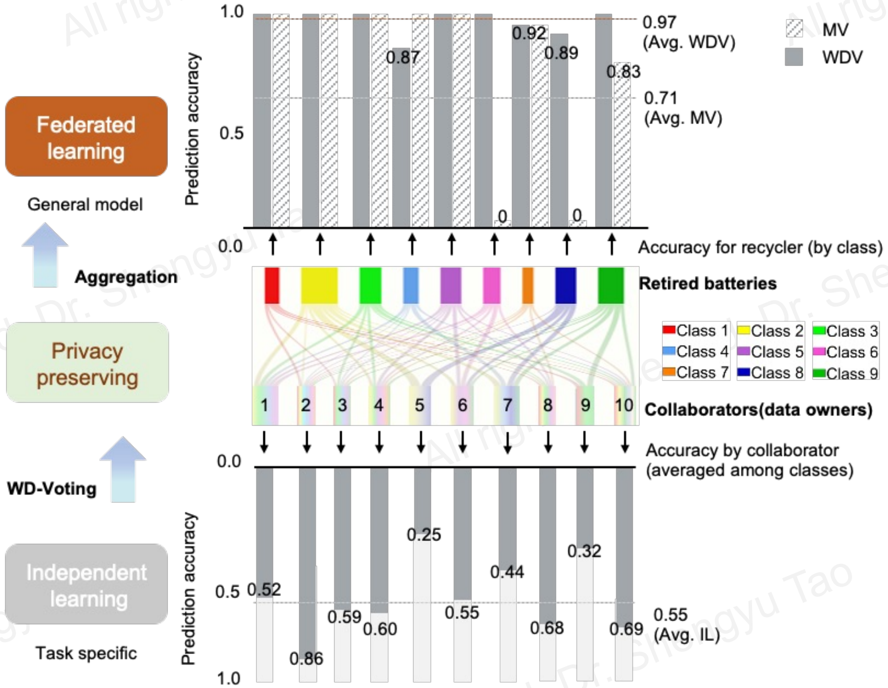


Figure 4.8 Class-wise (upper part) and client-wise (lower part) sorting accuracy.

Chapter 4: Collaborative Material Sorting

☐ Results

- ☐ **The model automatically finds the important features that relate to the dQ/dV peaks.**
- ☐ **The mixed battery materials are "sortable" only using F1 and F16 features (the dQ/dV peaks).**

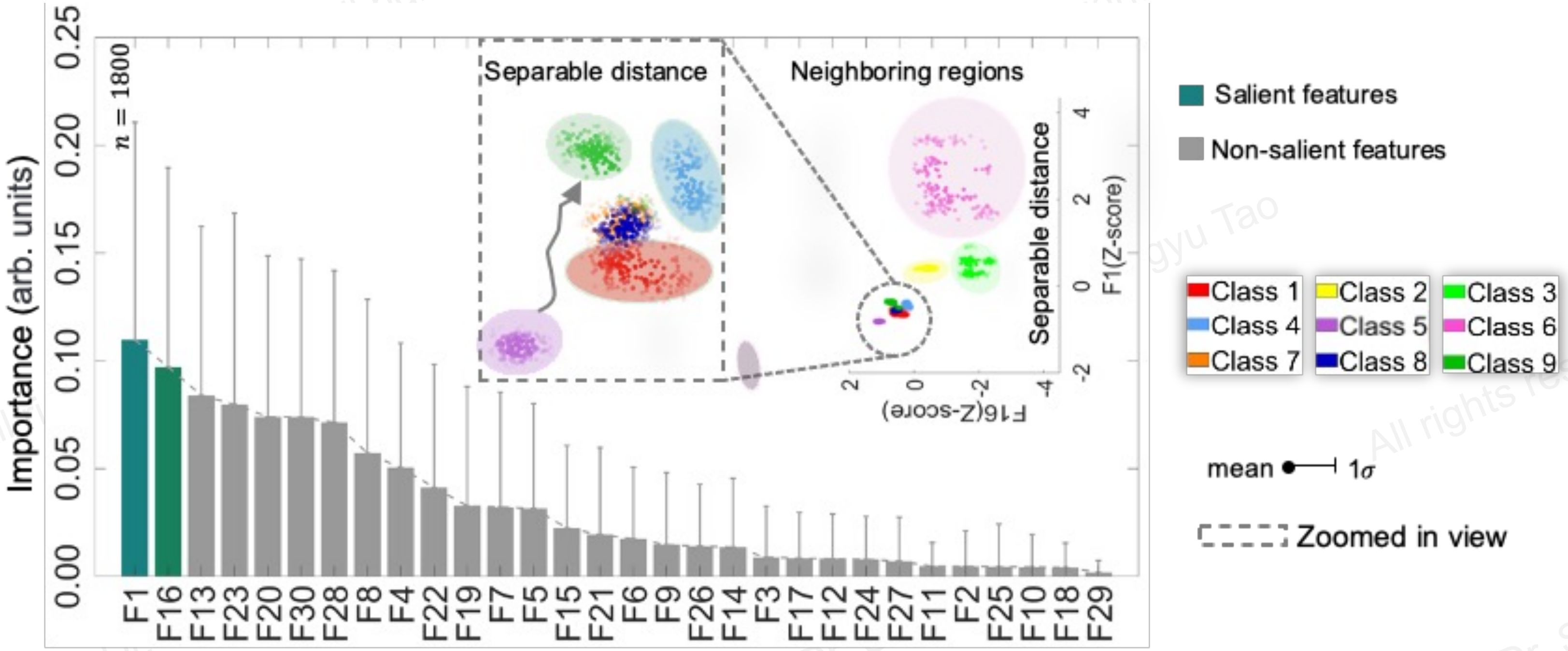
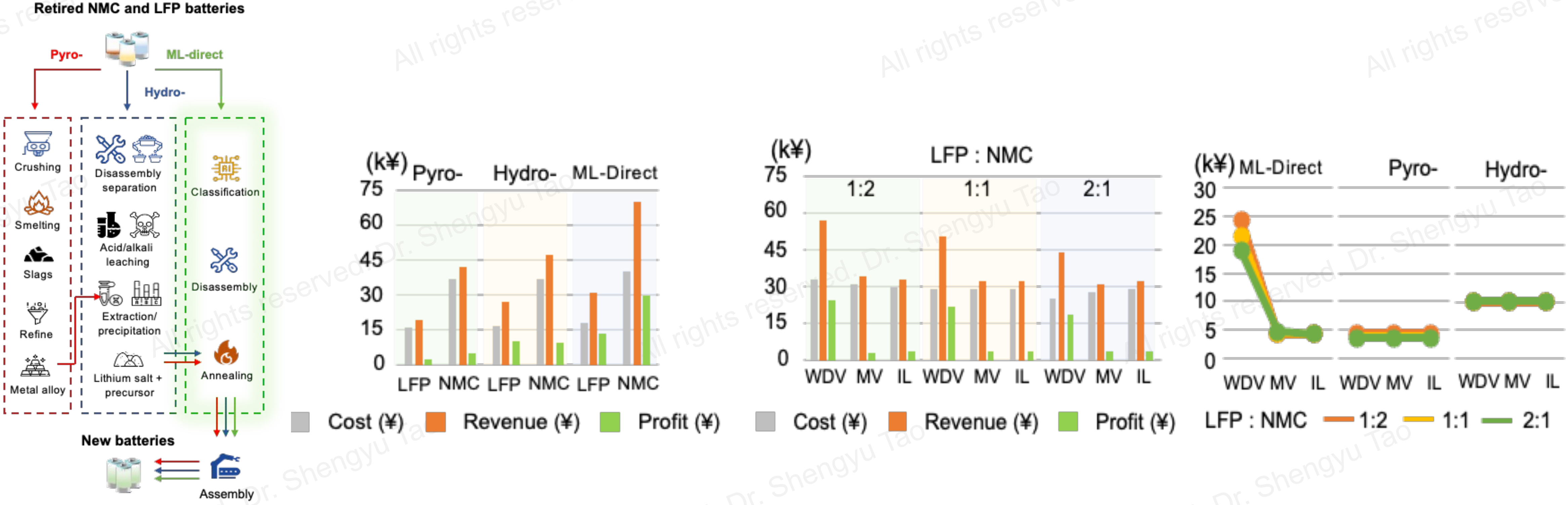


Figure 4.9 The feature importance analysis.

Chapter 4: Collaborative Material Sorting

☐ Results

☐ **The profit of direct recycling after material sorting increased over 60% averaged over different material mix rate.**



NMC: Nickel Manganese Cobalt Oxide
LFP: Lithium Iron Phosphate
ML: Machine Learning

Figure 4.10 The sustainability analysis.

Chapter 5:
Unified Diagnostics and Prognostics

Chapter 5: Unified Diagnostics and Prognostics

☐ Research Question

☐ **How to ensure muti-condition, multi-task, multi-metric battery lifecycle management using unified model?**

- Research gaps

- Existing data testing-based generation methods are costly and time-consuming
- Existing prediction methods are vulnerable to varying operation conditions and model parameters

- Research gaps

- Existing data generation cannot exhaust case-specific operation conditions
- Existing prediction methods are barely informed of diversified degradation mechanisms

- Solutions:

- Collaborate to maximize existing data source
  ->mitigate data restrictions
- Generate new condition-constrained data to ensure robustness
  ->increase data generalizability

- Solutions:

- Integrate condition information into adaptable diagnostics & prognosis
- Integrate physics (degradation mechanism) into battery diagnostics & prognosis

🟨 Research gaps
⬜ This work

Chapter 5: Unified Diagnostics and Prognostics

☐ Overview

☐ **Motivation**

- ☐ **The physical measurements are inconsistent.**
- ☐ **The diagnostics and prognostics tasks are different.**
- ☐ **The model parameters are constantly changing but the model is hardly adaptive.**

☐ **Contribution**

- ☐ **CORAL transformation and neural network integration in a unified form.**
- ☐ **CORAL works for multiple tasks (RUL, SOC, SOH) and physical measurements (long term cycling and short-term pulse test)**

Chapter 5: Unified Diagnostics and Prognostics

☐ Methodology Summary

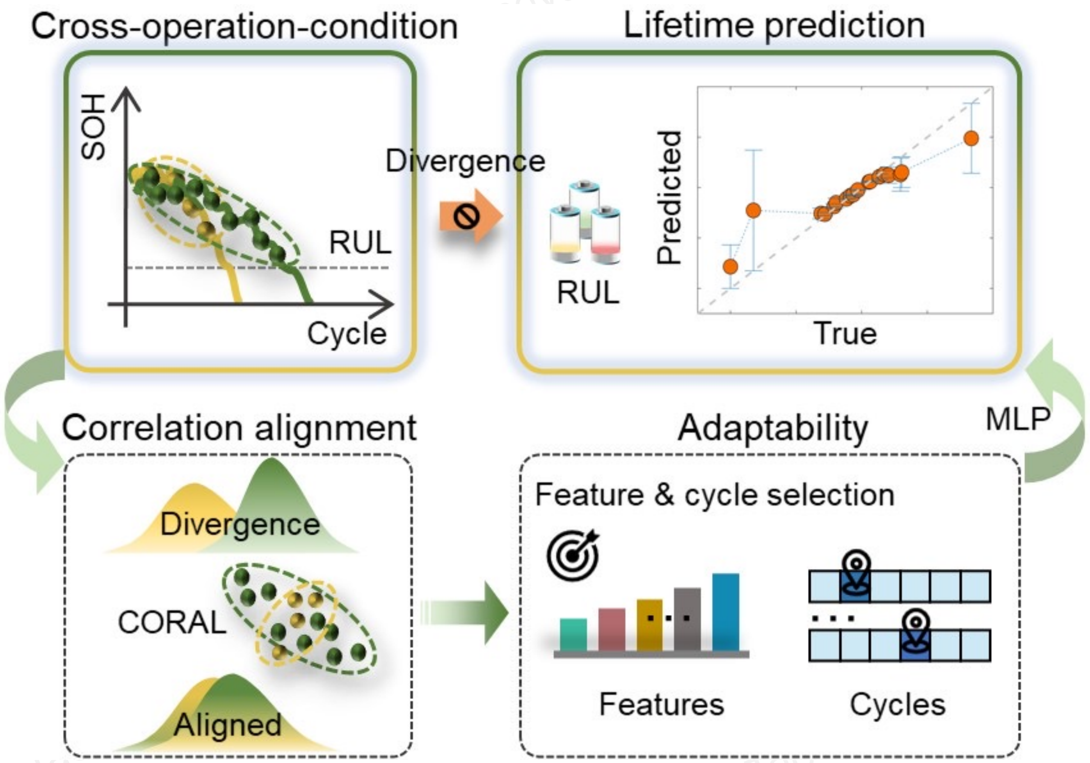


Figure 5.1 The feature importance analysis.

Chapter 5: Unified Diagnostics and Prognostics

☐ Methodology

☐ **The covariance matrix of source domain and target features:** $C_S = Cov(X_S) + \lambda I_p, C_T = Cov(X_T) + \lambda I_p$

where, $\mathrm{I_p}$ is an identity matrix. $\lambda > 0$ is the regularization term. Smaller λ corresponds to more alignment, while larger λ corresponds to less alignment. $Cov(\cdot)$ is the covariance operator.

☐ **The domain difference minimization:** $\min_{A} \ \|C_{\hat{S}} - C_T\|_{\mathrm{F}}^2 = \min_{A} \ \|A^T C_S A - C_T\|_{\mathrm{F}}^2$

where, $C_{\hat{S}}$ is the covariance of the CORAL-transformed source domain features, A is a linear transformation applied to the D_S features. $\|\cdot\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} \left|a_{ij}\right|^2}$

☐ Adaptability: $Adaptability = 1 - W_q(\cdot,\cdot) = \left(\underset{\gamma \in MP}{inf} \int_{\Omega\times\Omega} |X_{\hat{S}} - X_T|^q d\gamma(X_{\hat{S}}, X_T)\right)^{\frac{1}{q}}$

where, where γ refers to the transportation of arbitrary CORAL-transformed feature points in the fast-charging condition $X_{\hat{S}}$ to the extremely fast-charging condition feature X_T. Ω is the original feature space.

Chapter 5: Unified Diagnostics and Prognostics

☐ Methodology

☐ **The CORAL loss:** $l_{\mathrm{CORAL}} = \frac{1}{4k^2}\|C_s - C_t\|_F^2$

where, C_s and C_t are the covariance matrices of source domain and target domain features, $\|\cdot\|_F^2$ is the squared Frobenius norm. k is the dimensionality of the features, and thus, $k = 21$. The factor $\frac{1}{4k^2}$ normalizes the loss concerning the dimensionality of features, ensuring that the loss values remain comparable even for different feature dimensions.

☐ The loss function of multi-task estimation (e.g., SOC and SOH estimations):

$$L_{CORAL\_net} = \boldsymbol{\alpha}^{\boldsymbol{T}} \cdot (l_{MSE}^{SOC_{src}} + l_{MSE}^{SOC_{tgt}} + l_{MSE}^{RRC_{src}} + l_{MSE}^{RRC_{tgt}}) + \beta \cdot l_{\mathrm{CORAL}}$$

where, $\boldsymbol{\alpha}$ is a weighting vector for source domain and target domain tasks, including SOC prediction and relative remaining capacity prediction. β is a scalar value for the CORAL loss as a regularization that penalizes the model to converge to directions with large domain divergence. l_{MSE} refers to the mean square error.

Chapter 5: Unified Diagnostics and Prognostics

☐ Results

- **The feature engineering shows great domain divergence with different operation conditions.**

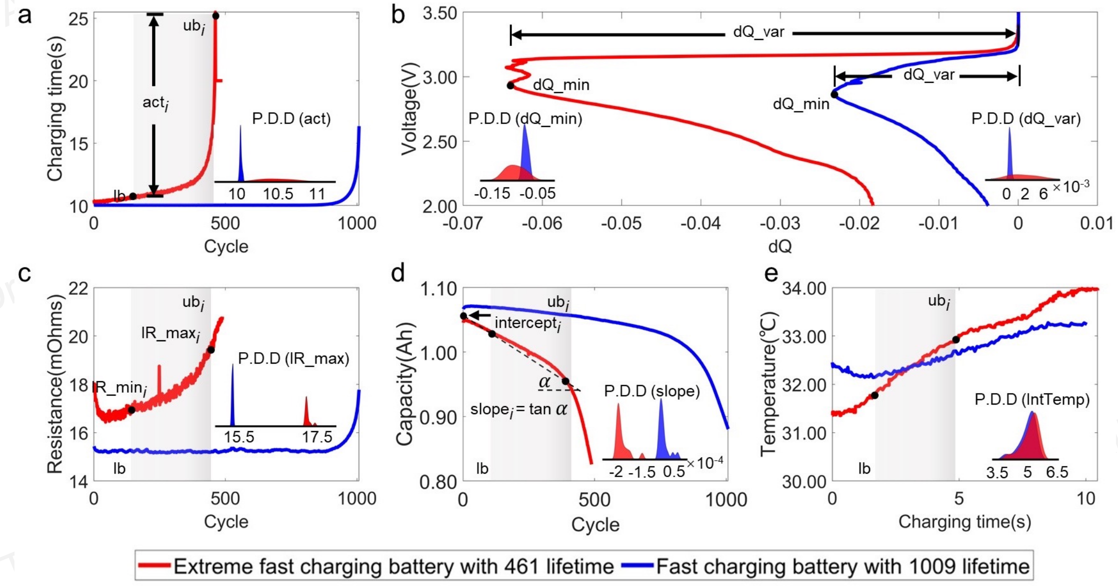
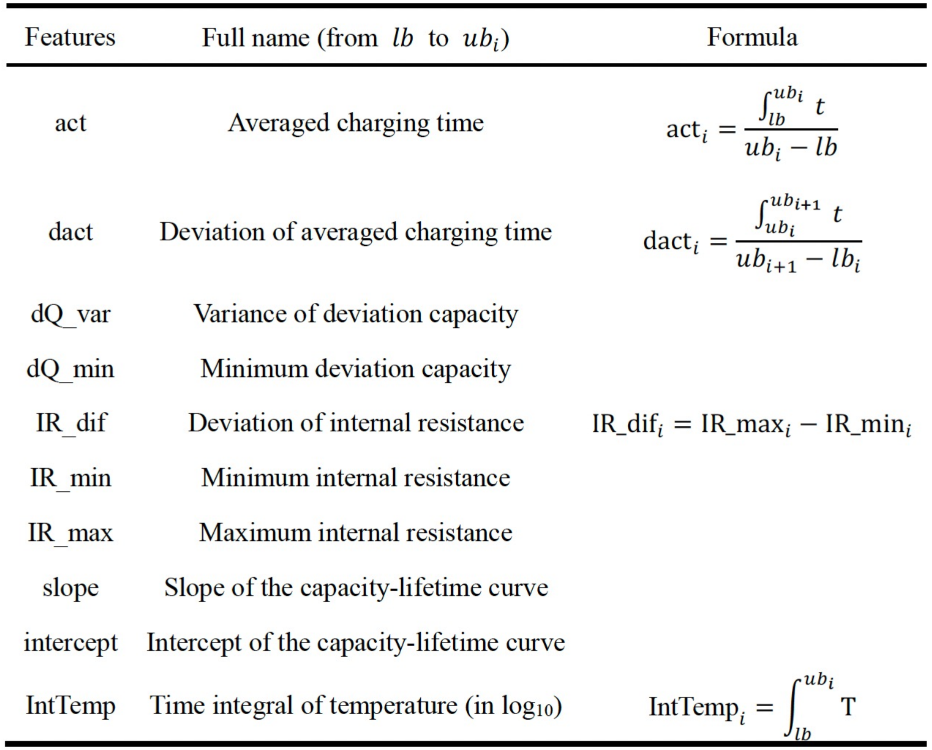


Figure 5.2 The feature engineering visualization.



Features	Full name (from lb to ub_i)	Formula
act	Averaged charging time	$\text{act}_i = \frac{\int_{lb}^{ub_i} t}{ub_i - lb}$
dact	Deviation of averaged charging time	$\text{dact}_i = \frac{\int_{ub_i}^{ub_{i+1}} t}{ub_{i+1} - lb_i}$
dQ_var	Variance of deviation capacity	
dQ_min	Minimum deviation capacity	
IR_dif	Deviation of internal resistance	$\text{IR\_dif}_i = \text{IR\_max}_i - \text{IR\_min}_i$
IR_min	Minimum internal resistance	
IR_max	Maximum internal resistance	
slope	Slope of the capacity-lifetime curve	
intercept	Intercept of the capacity-lifetime curve	
IntTemp	Time integral of temperature (in $\log_{10}$)	$\text{IntTemp}_i = \int_{lb}^{ub_i} \text{T}$

Figure 5.3 The feature engineering methods.

Chapter 5: Unified Diagnostics and Prognostics

☐ Results

- ☐ **CORAL-aided MLP reduced 98% RUL prediction error with an absolute error of 31.8 cycles.**
- ☐ **CORAL-aided predictions are more stable than the state-of-the-art fine-tuning methods.**

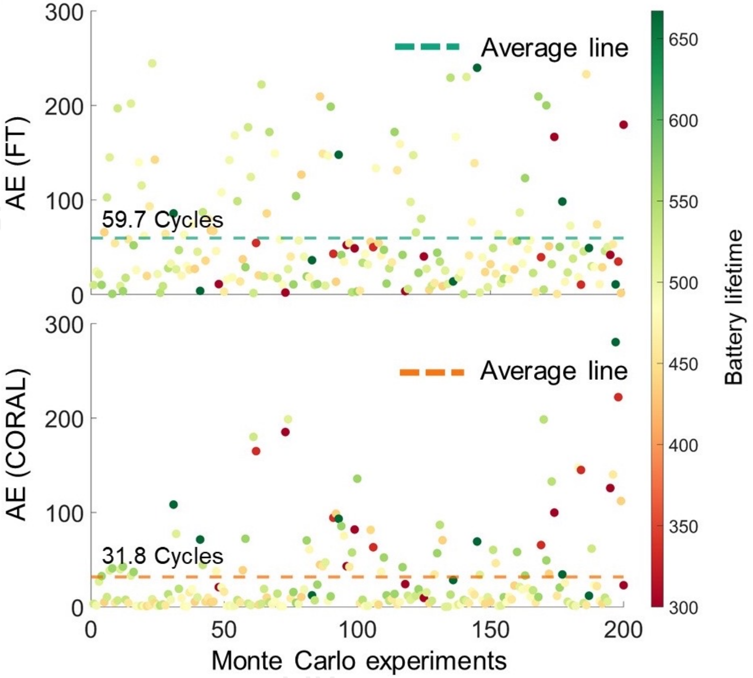


Figure 5.4 The lifetime prediction error.

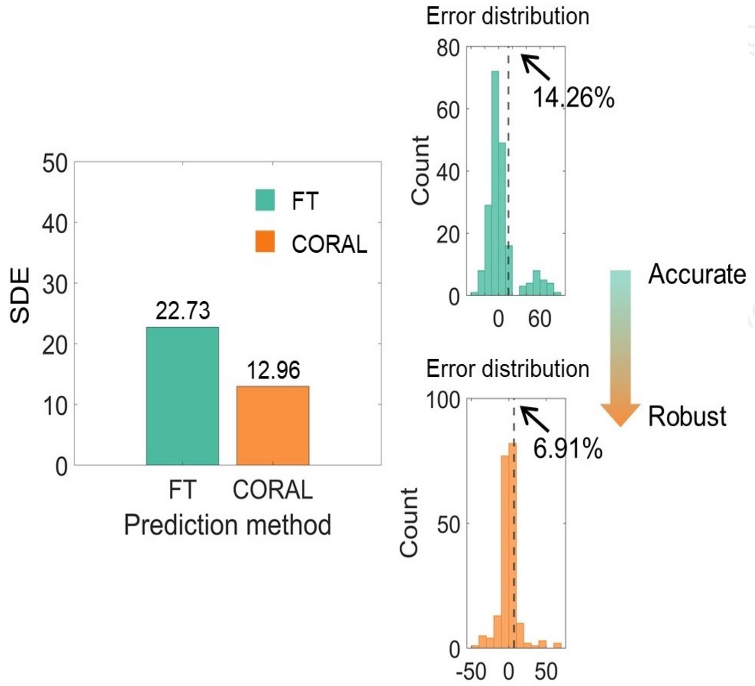


Figure 5.5 The lifetime prediction stability.

Chapter 5: Unified Diagnostics and Prognostics

☐ Results

- ☐ **The average error of 6.9% (31 cycles) shows stability with different regularization parameters**
- ☐ **CORAL can effectively align feature distributions by minimizing domain divergence.**

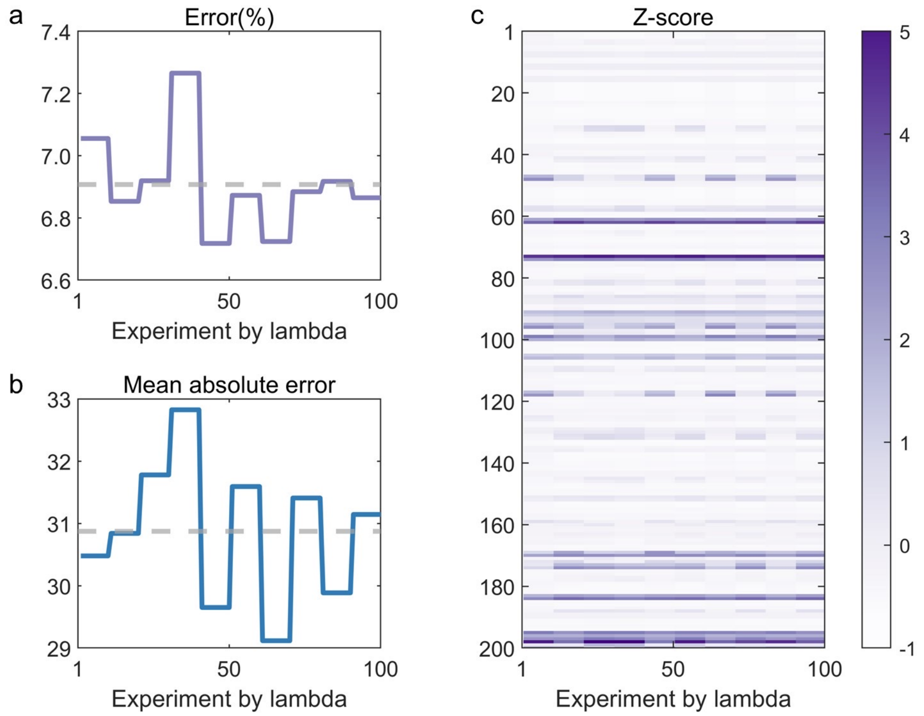


Figure 5.6 The lifetime prediction error.

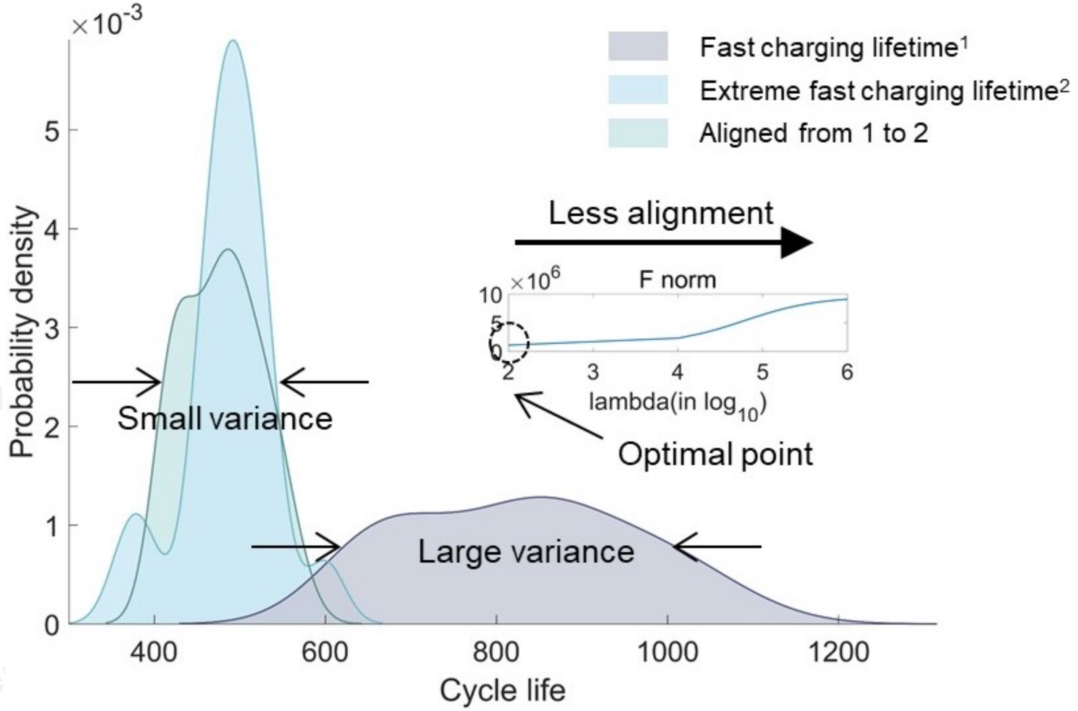


Figure 5.7 The effectiveness of feature alignment.

Chapter 5: Unified Diagnostics and Prognostics

☐ Results

- ☐ **The cycle-wise adaptability observes a “saturation” after 120 cycles. (averaged by features)**
- ☐ **The dQ_min feature has the best feature-wise adaptability. (averaged by cycles)**
- ☐ **The prediction accuracy is positively related to features adaptability regardless of feature types and cycle time.**

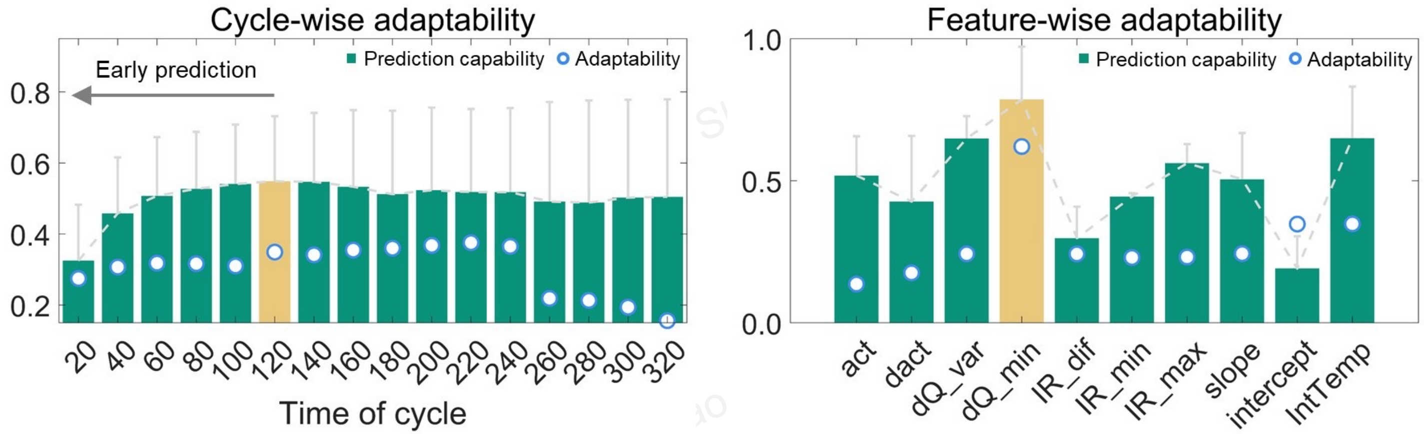


Figure 5.8 The adaptability evaluation.

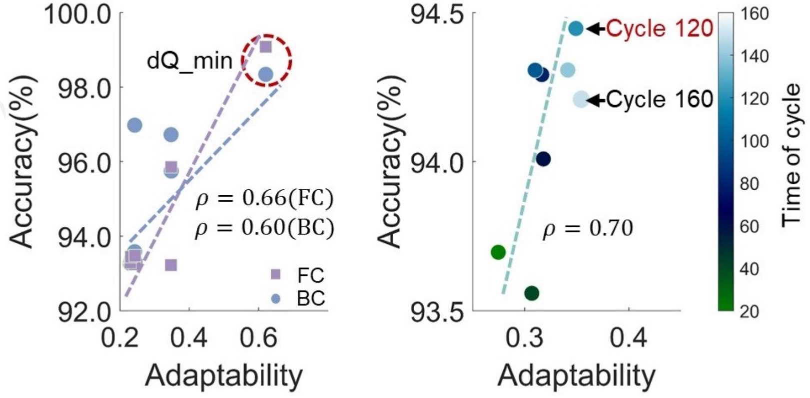


Figure 5.9 The effectiveness of feature alignment.

Chapter 5: Unified Diagnostics and Prognostics

☐ Results

- ☐ **CORAL can be used in different tasks (RUL, SOC, SOH) with different form (neural network loss)**
- ☐ **CORAL can be used in different physical measurement (long term charging and short term pulse test)**
- ☐ **CORAL outperforms existing methods under all data accesses.**

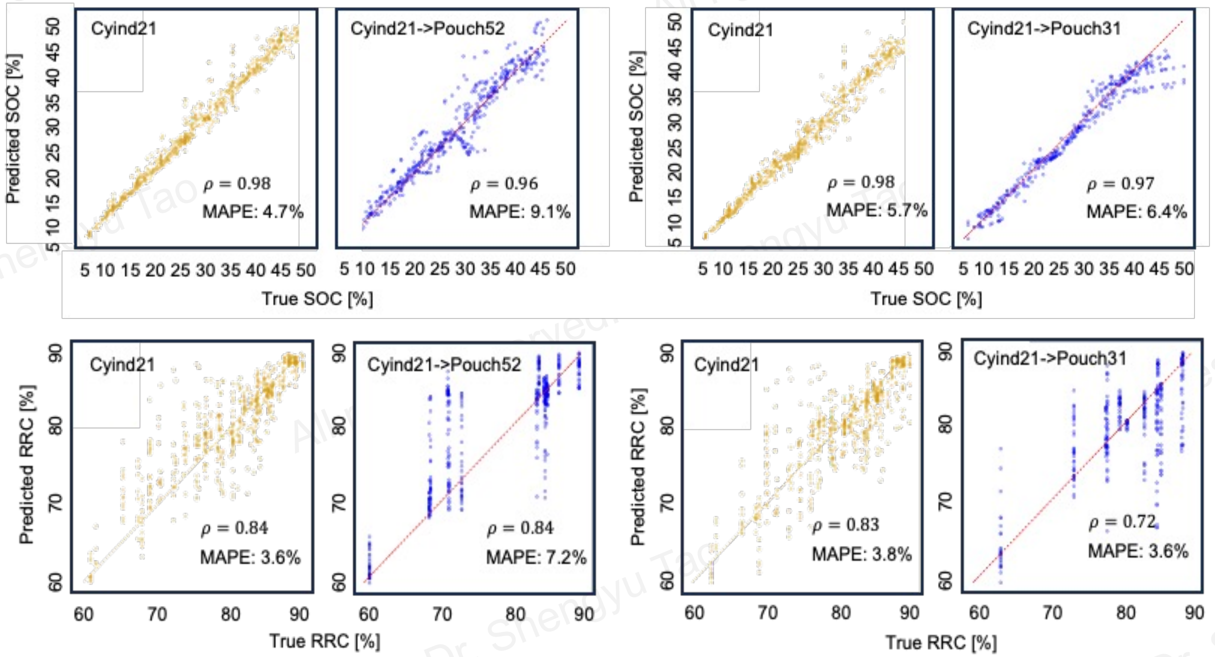


Figure 5.10 SOC and SOH estimation using CORAL

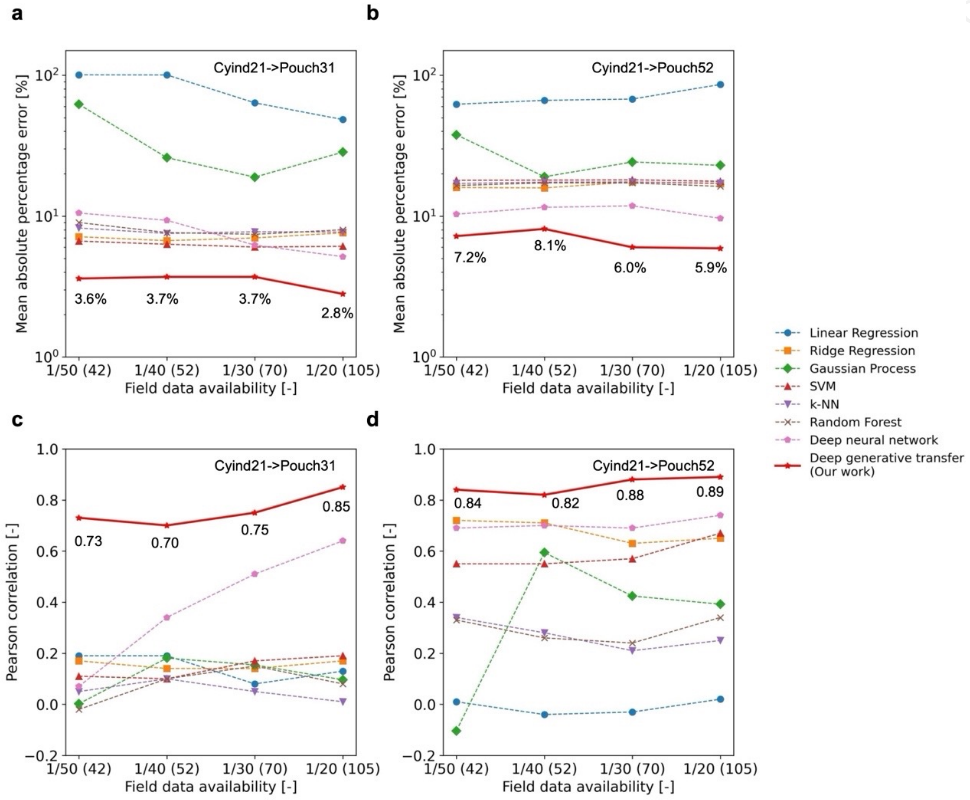


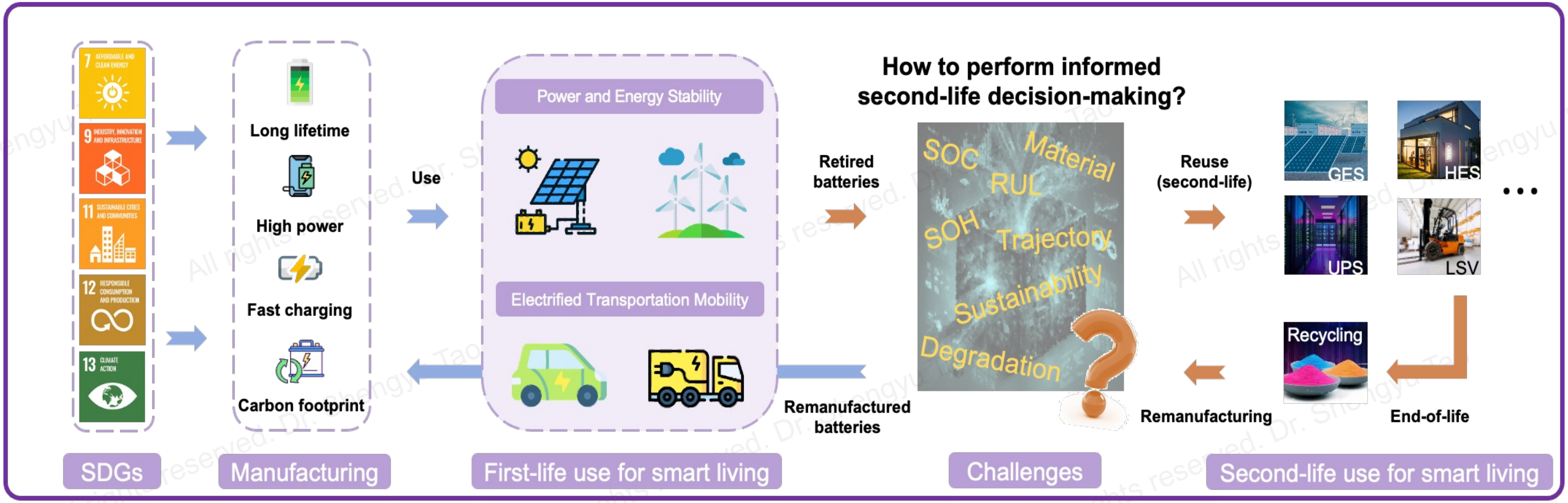
Figure 5.11 CORAL model benchmarking.

Chapter 6: Conclusions & Outlooks

Chapter 6: Conclusions & Outlooks

☐ Conclusions

- ☐ Machine learning has been seamlessly integrated into battery remanufacturing, reusing, and recycling by addressing data scarcity and heterogeneity challenges for maximized efficiency, safety, and sustainability.
- ☐ Machine learning answers whether, when and how to perform remanufacturing, reusing, and recycling.



Chapter 6: Conclusions & Outlooks

☐ Outlooks

- ☐ Non-invasive characterizations for prior knowledge about battery degradation, thus enabling effective integration of physics knowledge into data-driven models.
- ☐ Evaluating the informativeness of existing datasets to ensure meaningful data expansion with careful consideration of data sufficiency.
- ☐ Techno-economic analysis of data-driven models, from data curation, data storage, data processing, model training and model deployment.
- ☐ Large language models that streamline literature reviews, synthesis, reports, specification sheets, and extraction of technical knowledge.
- ☐ Model evaluation that follows industry-standard benchmarks and defines acceptable error margins.

Publications

Publications

First-author papers:

[1] **TAO S**, ZHANG M, ZHAO Z, et al. Non-destructive degradation pattern decoupling for early battery trajectory prediction via physics-informed learning [J]. *Energy & Environmental Science*, 2025, **18**(3): 1544-59. **IF=32**

[2] **TAO S**, MA R, ZHAO Z, et al. Generative learning assisted state-of-health estimation for sustainable battery recycling with random retirement conditions [J]. *Nature Communications*, 2024, **15**(1): 10154**. IF=14**

[3] **TAO S**, MA R, CHEN Y, et al. Rapid and sustainable battery health diagnosis for recycling pretreatment using fast pulse test and random forest machine learning [J]. *Journal of Power Sources*, 2024, **597**: 234156. **IF=9**

[4] **TAO S**, LIU H, SUN C, et al. Collaborative and privacy-preserving retired battery sorting for profitable direct recycling via federated machine learning [J]. *Nature Communications*, 2023, **14**(1): 8032. **IF=14**

[5] **TAO S**, SUN C, FU S, et al. Battery Cross-Operation-Condition Lifetime Prediction via Interpretable Feature Engineering Assisted Adaptive Machine Learning [J]. *ACS Energy Letters*, 2023, **8**(8): 3269-79. **IF=22**

[6] **TAO S**, GUO R, Jaewoong L, et al. Immediate remaining capacity estimation of heterogeneous second-life lithium-ion batteries via deep generative transfer learning [J]. *Energy & Environmental Science*, 2025,**18**, 7413-7426. **IF=32**

Publications

Co-first-author papers:

[7] HUANG X, **TAO S**, LIANG C, et al. Robust and generalizable lithium-ion battery health estimation using multi-scale field data decomposition and fusion [J]. *Journal of Power Sourc*es, 2025, **642**: 236939.
[8] LIANG C, **TAO S**, HUANG X, et al. Stochastic state of health estimation for lithium-ion batteries with automated feature fusion using quantum convolutional neural network [J]. *Journal of Energy Chemistry*, 2025, **106**: 205-19.
[9] MA R, **TAO S**, SUN X, et al. Pathway decisions for reuse and recycling of retired lithium-ion batteries considering economic and environmental functions [J]. *Nature Communications*, 2024, **15**(1): 7641.
[10] FU S, **TAO S**, FAN H, et al. Data-driven capacity estimation for lithium-ion batteries with feature matching based transfer learning method [J]. *Applied Energy*, 2024, **353**: 121991.
[11] LIU X, **TAO S**, FU S, et al. Binary multi-frequency signal for accurate and rapid electrochemical impedance spectroscopy acquisition in lithium-ion batteries [J]. *Applied Energy*, 2024, **364**: 123221.
[12] TALIHATI B, **TAO S**, FU S, et al. Energy storage sharing in residential communities with controllable loads for enhanced operational efficiency and profitability [J]. *Applied Energy*, 2024, **373**: 123880.
[13] HE K, **TAO S**, FU S, et al. A Novel Quick Screening Method for the Second Usage of Parallel-connected Lithium-ion Cells Based on the Current Distribution [J]. *Journal of The Electrochemical Society*, 2023, **170**(3): 030514.
[14] On the way…

References

☐ Reference List

1. Severson, K.A., Attia, P.M., Jin, N. *et al.* Data-driven prediction of battery cycle life before capacity degradation. *Nat Energy* **4**, 383–391 (2019).
2. Attia, P.M., Grover, A., Jin, N. et al. Closed-loop optimization of fast-charging protocols for batteries with machine learning. Nature 578, 397–402 (2020).
3. G. Ma, S. Xu, B. Jiang, et al. Real-time personalized health status prediction of lithium-ion batteries using deep transfer learning. Energy & Environmental Science 2022 Vol. 15 Issue 10 Pages 4083-4094.
4. Zhu, J., Wang, Y., Huang, Y. et al. Data-driven capacity estimation of commercial lithium-ion batteries from voltage relaxation. Nat Commun 13, 2261 (2022).
5. dos Reis, G., Strange, C., Yadav, M. & Li, S. Lithium-ion battery data and where to find it. *Energy and AI* **5**, 100081 (2021).
6. Hu, X., Li, S. & Peng, H. A comparative study of equivalent circuit models for Li-ion batteries. *Journal of Power Sources* **198**, 359-367 (2012). Wong, K. L., Chou, K. S., Tse, R., Tang, S.-K. & Pau, G. A Novel Fusion Approach Consisting of GAN and State-of-Charge Estimator for Synthetic Battery Operation Data Generation. *Electronics* **12**, 657 (2023).
7. Naaz, F., Herle, A., Channegowda, J., Raj, A. & Lakshminarayanan, M. A generative adversarial network-based synthetic data augmentation technique for battery condition evaluation. *International Journal of Energy Research* **45**, 19120-19135 (2021).
8. Tang, X. *et al.* Recovering large-scale battery aging dataset with machine learning. *Patterns* **2**, 100302 (2021).
9. Jiang, L, *et al*. Generating Comprehensive Lithium Battery Charging Data with Generative AI.
10. Che, Y., Deng, Z., Lin, X., Hu, L. & Hu, X. Predictive Battery Health Management With Transfer Learning and Online Model Correction. *IEEE Transactions on Vehicular Technology* **70**, 1269-1277 (2021).
11. Deng, Z., Lin, X., Cai, J. & Hu, X. Battery health estimation with degradation pattern recognition and transfer learning. *Journal of Power Sources* **525**, 231027 (2022).
12. Su, S. *et al.* A Hybrid Battery Equivalent Circuit Model, Deep Learning, and Transfer Learning for Battery State Monitoring. *IEEE Transactions on Transportation Electrification* **9**, 1113-1127 (2023).
13. Li, Y., Li, K., Liu, X., Wang, Y. & Zhang, L. Lithium-ion battery capacity estimation — A pruned convolutional neural network approach assisted with transfer learning. *Applied Energy* **285**, 116410 (2021).
14. Shen, S., Sadoughi, M., Li, M., Wang, Z. & Hu, C. Deep convolutional neural networks with ensemble learning and transfer learning for capacity estimation of lithium-ion batteries. *Applied Energy* **260**, 114296 (2020).
15. Fu, S. *et al.* Data-driven capacity estimation for lithium-ion batteries with feature matching based transfer learning method. *Applied Energy* **353**, 121991 (2024).

References

☐ Reference List

17. Tao, S. *et al.* Battery Cross-Operation-Condition Lifetime Prediction via Interpretable Feature Engineering Assisted Adaptive Machine Learning. *ACS Energy Letters*, 3269-3279 (2023).
18. Che, Y., Zheng, Y., Onori, S., Hu, X. & Teodorescu, R. Increasing generalization capability of battery health estimation using continual learning. *Cell Reports Physical Science* **4**, 101743 (2023).
19. Lee, M. Bilevel-optimized continual learning for predicting capacity degradation of lithium-ion batteries. *Journal of Energy Storage* **86**, 111187 (2024).
20. Eaty, N. D. K. M. & Bagade, P. Digital twin for electric vehicle battery management with incremental learning. *Expert Systems with Applications* **229**, 120444 (2023).
21. Liu, X. *et al.* Fast sorting method of retired batteries based on multi-feature extraction from partial charging segment. *Applied Energy* **351**, 121930 (2023).
22. Qian, C. *et al.* Convolutional neural network based capacity estimation using random segments of the charging curves for lithium-ion batteries. *Energy* **227**, 120333 (2021).
23. Deng, Z.,. Data-Driven Battery State of Health Estimation Based on Random Partial Charging Data. *IEEE Transactions on Power Electronics* **37**, 5021-5031 (2022).
24. Yao, J. & Han, T. Data-driven lithium-ion batteries capacity estimation based on deep transfer learning using partial segment of charging/discharging data. *Energy* **271**, 127033 (2023).
25. Deng, Z. *et al.* Battery health evaluation using a short random segment of constant current charging. *iScience* **25**, 104260 (2022).
26. Zhang, C. *et al.* Enhancing state of charge and state of energy estimation in Lithium-ion batteries based on a TimesNet model with Gaussian data augmentation and error correction. *Applied Energy* **359**, 122669 (2024).
27. Tao, S., Liu, H., Sun, C. *et al.* Collaborative and privacy-preserving retired battery sorting for profitable direct recycling via federated machine learning. *Nat Commun* **14**, 8032 (2023).
28. Lin, C., Xu, J., Hou, J., Liang, Y. & Mei, X. Ensemble Method With Heterogeneous Models for Battery State-of-Health Estimation. *IEEE Transactions on Industrial Informatics* **19**, 10160-10169 (2023).
29. Wu, J., Chen, J., Feng, X., Xiang, H. & Zhu, Q. State of health estimation of lithium-ion batteries using Autoencoders and Ensemble Learning. *Journal of Energy Storage* **55**, 105708 (2022).
30. Xiang, H., Wang, Y., Zhang, X. & Chen, Z. Two-level Battery Health Diagnosis using Encoder-decoder Framework and Gaussian Mixture Ensemble Learning Based on Relaxation Voltage. *IEEE Transactions on Transportation Electrification*, 1-1 (2023).
31. Meng, J. *et al.* An optimized ensemble learning framework for lithium-ion Battery State of Health estimation in energy storage system. *Energy* **206**, 118140 (2020).

References

☐ Reference List

32. Gou, B., Xu, Y. & Feng, X. An Ensemble Learning-Based Data-Driven Method for Online State-of-Health Estimation of Lithium-Ion Batteries. *IEEE Transactions on Transportation Electrification* **7**, 422-436 (2021).
33. Guo, N. *et al.* Semi-supervised learning for explainable few-shot battery lifetime prediction. *Joule* (2024).
34. Kröger, T., Belnarsch, A., Bilfinger, P., Ratzke, W. & Lienkamp, M. Collaborative training of deep neural networks for the lithium-ion battery aging prediction with federated learning. *eTransportation* **18**, 100294 (2023).
35. Zhao, S. *et al.* Potential to transform words to watts with large language models in battery research. *Cell Reports Physical Science* **5**, 101844 (2024).

PhD Work Summary at Tsinghua & Berkeley

Machine Learning-Assisted Sustainable Remanufacturing, Reusing and Recycling for Lithium-ion Batteries

Dr. Shengyu Tao
Institute of Data and Information, Tsinghua University
Acknowledgments to Dr. Xuan Zhang, Dr. Guangmin Zhou and Dr. Scott Moura
2025/9/16

Appendix A

☐ Multi-step fast charging

Charging details	Time duration (min)	SOC
Step1: Rest	30.00	
Step2: $0.33C_{RPT}$ CC to 2.5V	-	
Step3: Rest	30.00	
Step4: $0.33C_{RPT}$ CC to U1	14.54	+8.0%
Step5: $3.00C_{RPT}$ CC to U2	2.40	+12.0%
Step6: $2.90C_{RPT}$ CC to U3	2.07	+10.0%
Step7: $2.80C_{RPT}$ CC to U4	2.14	+10.0%
Step8: $2.40C_{RPT}$ CC to U5	2.50	+10.0%
Step9: $2.00C_{RPT}$ CC to U6	3.00	+11.1%
Step10: $1.80C_{RPT}$ CC to U7	3.33	+10.0%
Step11: $1.40C_{RPT}$ CC to U8	4.29	+10.0%
Step12: $0.33C_{RPT}$ CC to U9	28.93	+15.9%
Step13: Rest	120.00	Summation: 97%
Step14: $1C_{RPT}$ CC to (U10)	56.40	-94%
Step15: Rest	60.00	
Step16: Repeat	Steps 3 to 14 are repeated 3 times. Mean values of (U1-U9) are taken as cut-off voltages for subsequent cycling.	

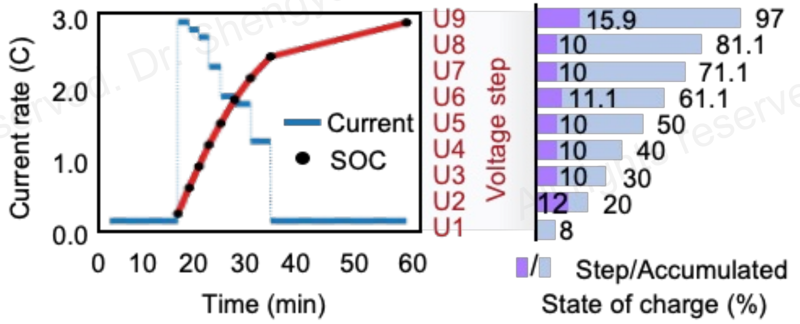


Figure A.1 The multi-step charging profile.

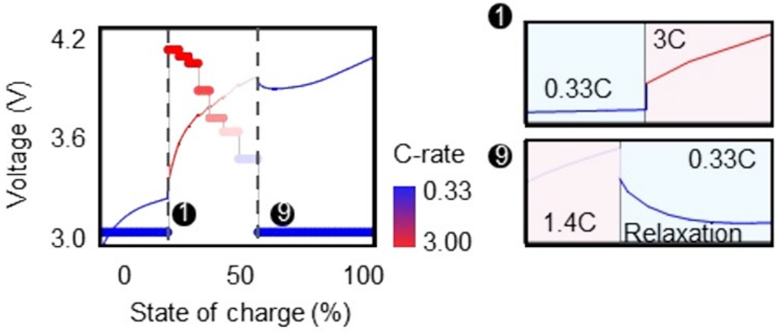


Figure A.2 The high current and low current region.

Appendix B

☐ The concept of thermodynamics and kinetics

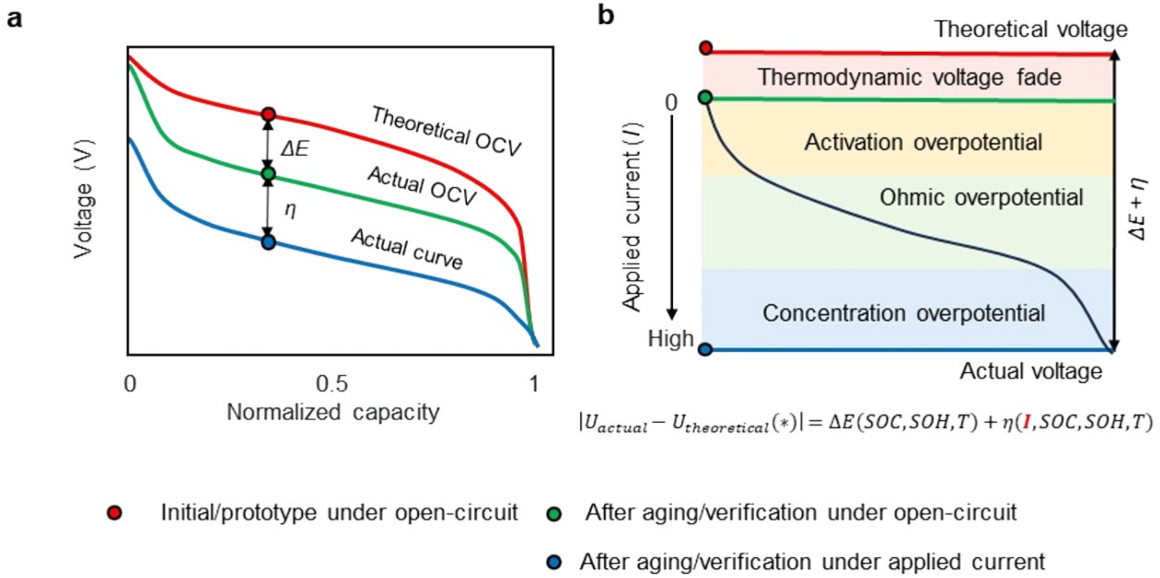


Figure B.1 The concept of thermodynamics and kinetics.

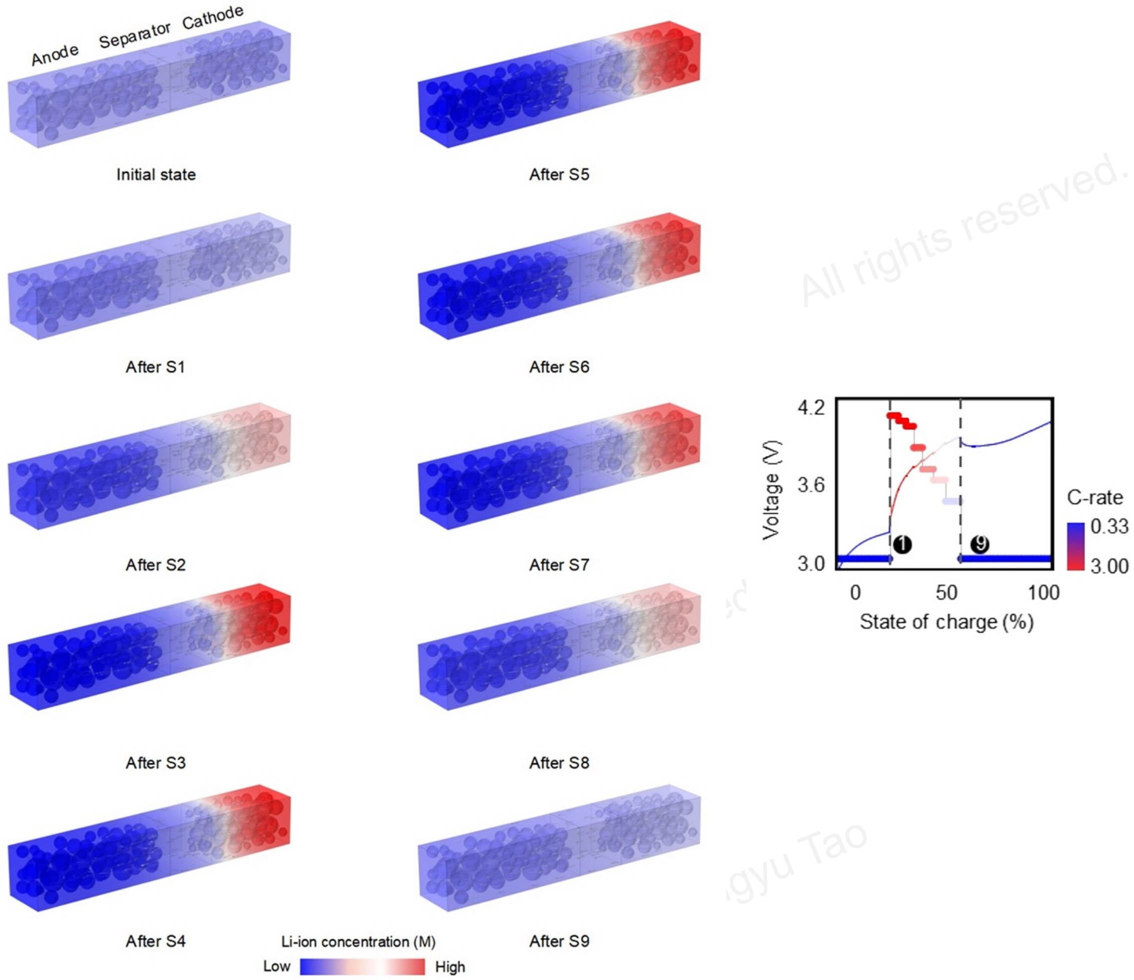
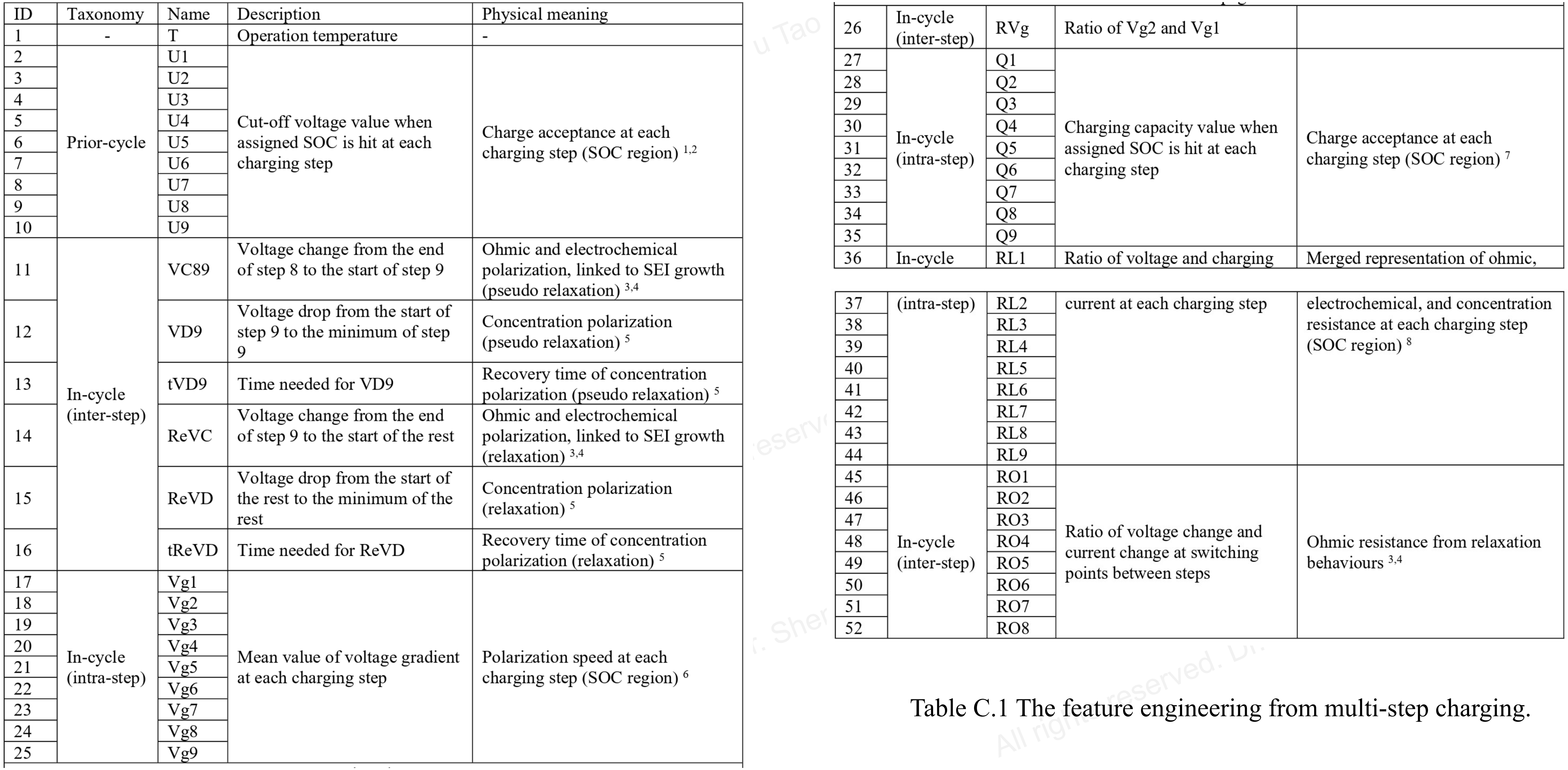


Figure B.2 The visualization of thermodynamics and kinetics.

Appendix C

☐ Thermodynamic and kinetic features



ID	Taxonomy	Name	Description	Physical meaning
1	-	T	Operation temperature	-
2	Prior-cycle	U1	Cut-off voltage value when assigned SOC is hit at each charging step	Charge acceptance at each charging step (SOC region) [1,2]
3		U2		
4		U3		
5		U4		
6		U5		
7		U6		
8		U7		
9		U8		
10		U9		
11	In-cycle (inter-step)	VC89	Voltage change from the end of step 8 to the start of step 9	Ohmic and electrochemical polarization, linked to SEI growth (pseudo relaxation) [3,4]
12		VD9	Voltage drop from the start of step 9 to the minimum of step 9	Concentration polarization (pseudo relaxation) [5]
13		tVD9	Time needed for VD9	Recovery time of concentration polarization (pseudo relaxation) [5]
14		ReVC	Voltage change from the end of step 9 to the start of the rest	Ohmic and electrochemical polarization, linked to SEI growth (relaxation) [3,4]
15		ReVD	Voltage drop from the start of the rest to the minimum of the rest	Concentration polarization (relaxation) [5]
16		tReVD	Time needed for ReVD	Recovery time of concentration polarization (relaxation) [5]
17	In-cycle (intra-step)	Vg1	Mean value of voltage gradient at each charging step	Polarization speed at each charging step (SOC region) [6]
18		Vg2		
19		Vg3		
20		Vg4		
21		Vg5		
22		Vg6		
23		Vg7		
24		Vg8		
25		Vg9		
26	In-cycle (inter-step)	RVg	Ratio of Vg2 and Vg1	
27	In-cycle (intra-step)	Q1	Charging capacity value when assigned SOC is hit at each charging step	Charge acceptance at each charging step (SOC region) [7]
28		Q2		
29		Q3		
30		Q4		
31		Q5		
32		Q6		
33		Q7		
34		Q8		
35		Q9		
36	In-cycle (intra-step)	RL1	Ratio of voltage and charging current at each charging step	Merged representation of ohmic, electrochemical, and concentration resistance at each charging step (SOC region) [8]
37		RL2		
38		RL3		
39		RL4		
40		RL5		
41		RL6		
42		RL7		
43		RL8		
44		RL9		
45	In-cycle (inter-step)	RO1	Ratio of voltage change and current change at switching points between steps	Ohmic resistance from relaxation behaviours [3,4]
46		RO2		
47		RO3		
48		RO4		
49		RO5		
50		RO6		
51		RO7		
52		RO8		

Table C.1 The feature engineering from multi-step charging.

Appendix D

☐ Pulse test

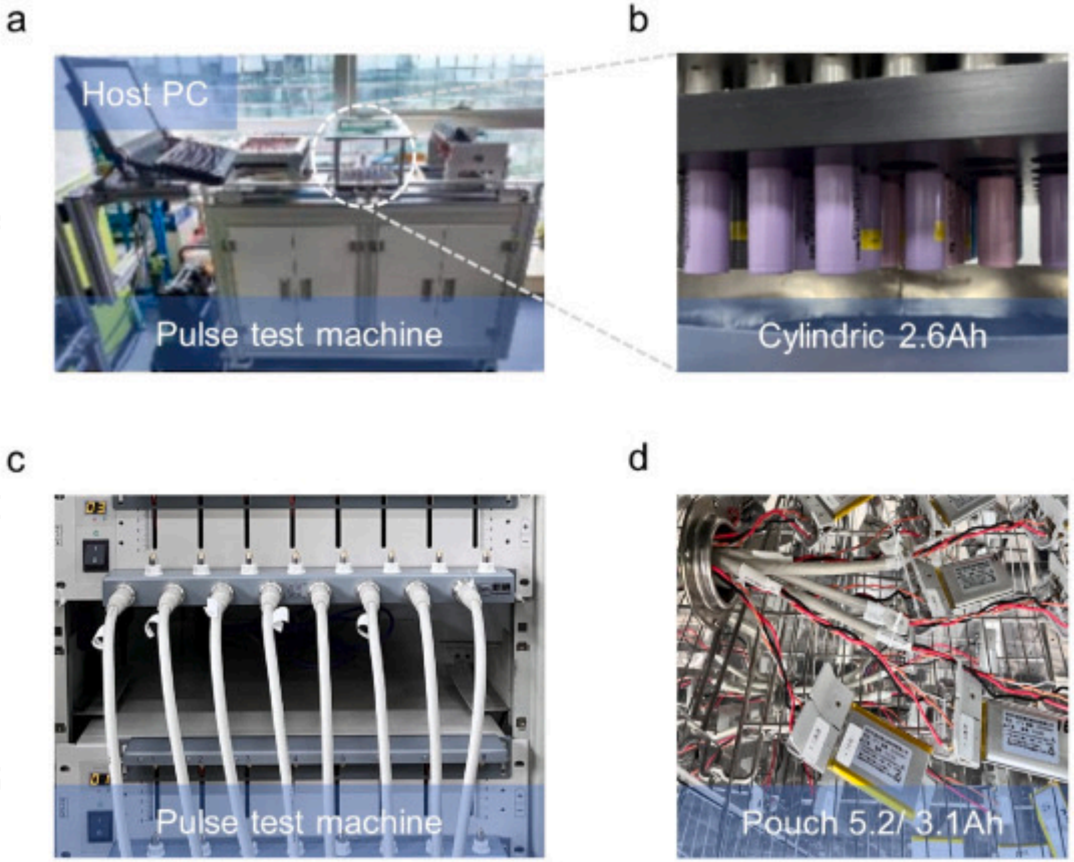


Figure D.1 The part of the pulse test equipment.

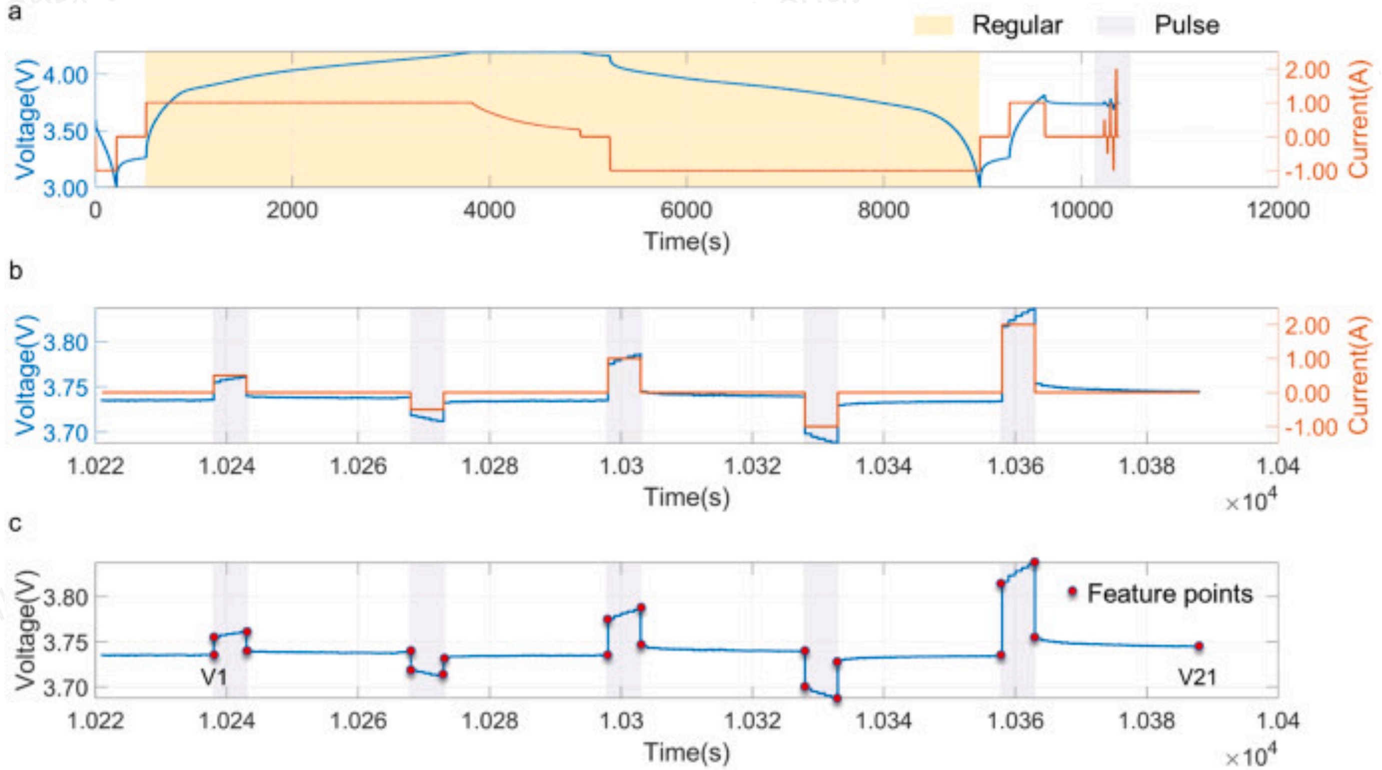


Figure D.2 The schematics of pulse test and feature engineering.

PhD Work Summary at Tsinghua & Berkeley

Machine Learning-Assisted Sustainable Remanufacturing, Reusing and Recycling for Lithium-ion Batteries

Dr. Shengyu Tao

Thanks for your time and comments.

Collaborations and discussions are always welcome.

Contact: terencetaotbsi@gmail.com

Scan the QR code