Shared Representation Learning for Generalizable SOH Estimation Across Multiple Battery Configurations

Shunyu Wu*, Zhuomin Chen*, Bingxin Lin, Haozheng Ye, Jiahui Zhou, Yanran Zhao, Dan Li†, Jian Lou

School of Software Engineering, Sun Yat-Sen University, Zhuhai 519080, China

School of the Art Institute of Chicago
{*wushy88, chenzhm39, linbx25, yehzh8, zhoujh99*}*@mail2.sysu.edu.cn*
yzhao57@artic.edu
{*lidan263, louj5*}*@mail.sysu.edu.cn*

ABSTRACT

Battery health monitoring is essential in applications such as electric vehicles and energy storage systems, where the lifespan and health state of batteries directly impact the safety and operational costs. However, existing works have demonstrated promising performance in predicting the state of health (SOH) of batteries within the same type under certain working conditions. However, batteries are produced with different types and work under different conditions in real applications. Existing methods fail to leverage the inherent correlations between related battery types and overlook the various working conditions, resulting in suboptimal robustness and prediction accuracy. To address this limitation, we propose SRSE: a novel Shared Representation learning framework that jointly learns shared representation (hidden knowledge) across multiple battery configurations for robust and generalized SOH Estimation. In particular, an adversarial training scheme is utilized to eliminate task-specific contamination in the shared feature space. SRSE captures both feature-level and logit-level shared knowledge and subsequently transfers it from the shared layer to task-specific layers, enhancing the adaptability and efficiency of each task. Extensive experiments on three large-scale battery health datasets demonstrate that our proposed method significantly improves SOH estimation performance across diverse battery types and operating conditions.

1. INTRODUCTION

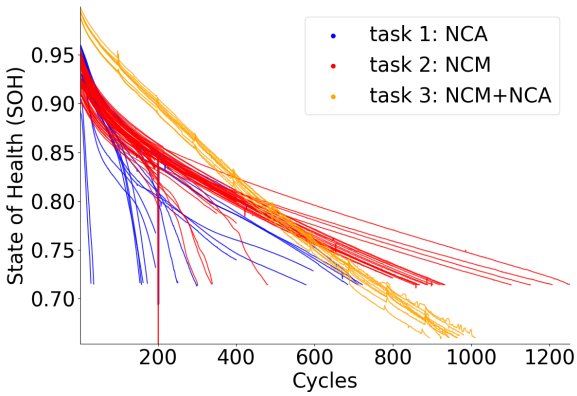
Battery health monitoring is a critical technology to ensure the safety and reliability of batteries, which are widely used in electric vehicles, energy storage systems, and portable electronic devices. Key indicators such as State of Health (SOH) and Remaining Useful Life (RUL) are essential for evaluating battery performance and determining maintenance and replacement strategies. While SOH reflects the current capacity retention, RUL extends this by predicting end-of-life timelines(Wen et al., 2025; Hou, Wang, et al., 2025; Y. Wang, Wu, Li, Xie, & Chen, 2025; Hou, Ragab, et al., 2025), with both relying on understanding battery degradation patterns. However, battery health monitoring faces significant challenges, including the non-linear and time-varying nature of battery aging processes, as well as the difficulty in directly observing internal states. Current battery health monitoring methods can be broadly categorized into physics-based and data-driven approaches(F. Wang, Zhai, Zhao, Di, & Chen, 2024). Many scholars estimate SOH by establishing battery aging models(Baghdadi, Briat, Delétage, Gyan, & Vinassa, 2016; Dong & Wei, 2021; Lui et al., 2021), leveraging prior knowledge, but suffer from high computational complexity and limited practicality. And data-driven methods(Attia et al., 2020; Rauf, Khalid, & Arshad, 2022) offer better real-time performance but lack interpretability and struggle to capture microscopic battery dynamics. Recent efforts to integrate physical knowledge with data-driven methods have shown promise, but limitations remain. Most existing approaches operate in a single-task learning paradigm, failing to leverage the inherent correlations between related tasks. By isolating tasks, those methods miss opportunities to exploit shared knowledge and dependencies, which could enhance the model's ability to generalize and improve prediction accuracy across multiple tasks. This highlights the need for developing a unified framework that can effectively integrate and utilize the interconnected nature of battery health monitoring tasks.

Multitask Learning (MTL) is a machine learning paradigm that improves model generalization and efficiency by sharing

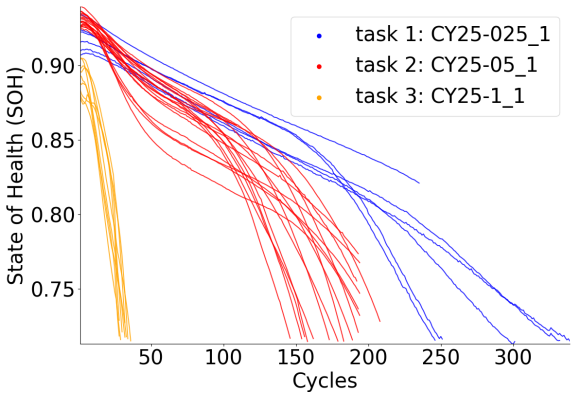
Shunyu Wu et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 United States License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

[1] *These authors contributed equally to this work.

[2] †Corresponding author.



(a) SOH degradation curves for batteries with different materials and models.



(b) SOH degradation curves for the same type of battery under different external conditions (charging rates).

Figure 1. Visualization of domain shifts in battery health monitoring across different battery types and operation conditions. (a) Differences in SOH degradation across battery types highlight the challenge of generalization in cross-type estimation tasks. (b) Even within the same battery type, external operation conditions such as charging rates induce distribution shifts, further complicating the problem for cross-condition generalization.

knowledge across multiple related tasks. Compared to single-task learning, sharing common knowledge across multiple learning tasks reduces overfitting, enhances data utilization, and improves the model's ability to handle complex tasks. MTL has achieved significant success in various domains, including computer vision(Heuer, Mantowsky, Bukhari, & Schneider, 2021), natural language processing(Chen, Zhang, & Yang, 2024), and healthcare(He et al., 2020). It has also attracted attention in battery health monitoring. For example, Zhang et al.(S. Zhang, Liu, Xu, & Su, 2024) proposed a physics-informed hybrid multitask learning (PIHMTL) method, which simultaneously learns about battery health status and other related tasks to estimate the full lifecycle aging state of lithium-ion batteries more comprehensively. However, these MTL approaches still face challenges, such as the difficulty in effectively disentangling shared and task-specific features, which limits model performance.

Despite its potential in battery health monitoring, significant research gaps remain in proposing a tailored MTL framework with satisfactory overall performance. First, existing methods struggle to effectively disentangle shared and task-specific features, leading to feature redundancy and noise interference in the shared feature space, which significantly degrades model performance. Second, task interference is a critical issue, where certain tasks may negatively impact the learning of others, especially when there are substantial differences between tasks. This interference often results in suboptimal performance for some tasks while improving others, creating an imbalance in task learning. Third, the efficiency of knowledge transfer in the existing MTL framework remains low. Figure 1 visualizes the domain shifts in battery health monitoring. Specifically, Figure 1a illustrates the differences in state of health (SOH) degradation curves across various battery types, highlighting the distribution shifts under varying battery conditions. Figure 1b demonstrates that even within the same type of battery, distribution shifts can occur under different external conditions, such as charging rates, further complicating the learning process. Consequently, due to the unresolved issue of domain shift, current methods still lack effective mechanisms to transfer learned knowledge from multiple tasks to new ones, thereby limiting the model's adaptability and generalization capabilities. These challenges highlight the need for more advanced knowledge-sharing mechanisms that can address feature redundancy, task interference, and knowledge distillation in battery health monitoring.

To address these gaps, this paper proposes a novel Shared Representation learning framework that integrates adversarial training, knowledge distillation, and domain adaptation techniques (Ben-David, Blitzer, Crammer, & Pereira, 2006; Ben-David et al., 2010), specifically designed for generalizable SOH Estimation across various battery configurations (SRSE). The framework introduces a joint feature purification mechanism that leverages adversarial training to ensure the shared feature space remains free from task-specific contamination. Specifically, we employ a gradient reversal layer (GRL) to enforce task-invariant feature learning, ensuring that the shared representation captures only common and task-invariant information. Additionally, we incorporate a knowledge distillation module that distills both feature-level and logit-level knowledge from the shared layer to the task-specific layers, enabling efficient cross-domain transfer. The feature-level fusion combines shared and task-specific features to produce more informative representations for each task, while the logit-level augmentation refines task-specific predictions by incorporating additional logit-level guidance

from the shared layer. This dual focus on feature purity and model efficiency addresses the limitations of existing methods and provides a robust solution for battery health monitoring tasks.

The contributions of this work are summarized as follows:

1. We propose a shared representation learning framework that integrates adversarial training, knowledge distillation, and domain adaptation techniques, specifically designed for generalizable SOH estimation across various battery configurations.
2. We introduce a joint feature purification mechanism that leverages adversarial training to ensure the shared feature space remains free from task-specific contamination.
3. We incorporate a knowledge distillation module that distills both feature-level and logit-level knowledge from the shared layer to the task-specific layers, enabling efficient cross-domain transfer.
4. We demonstrate the effectiveness of our proposed framework through extensive experiments on three large-scale battery health datasets, showing significant improvements in performance over existing methods.

2. RELATED WORK

Battery Health Monitoring is a critical task in battery management systems, aiming to accurately estimate the State of Health (SOH). Current battery health monitoring methods can be broadly categorized into physics-based and data-driven approaches. Physics-based methods rely on fundamental electrochemical and physical principles to model the behavior and degradation of lithium-ion batteries. Baghdadi et al.(Baghdadi et al., 2016) developed a lithium battery aging model based on Dakin's degradation approach, focusing on the calendar and power cycling aging behaviors of two distinct lithium battery chemistries. Similarly, Dong and Wei(Dong & Wei, 2021) developed an aging model that captures the coupled chemical and mechanical degradation mechanisms, including SEI layer formation and crack propagation, to accurately predict capacity loss under storage and cycling conditions. However, their performance is often limited by the accuracy of the assumed models, and those models may struggle to capture complex degradation patterns that are not fully described by existing physical models.

Then, with the rapid development of deep learning, data-driven methods gain significant attention due to their ability to capture complex patterns in battery data without relying on explicit physical models. For instance, Li et al. (Li, Lyv, Gao, Li, & Zhu, 2025) proposed a co-estimation framework combining semi-supervised learning (SSL) with Long Short-Term Memory (LSTM) networks to reduce the dependency on labeled data. Lu et al. (Lu, Xiong, Tian, Wang, & Sun, 2023) designed a deep learning framework to estimate battery SOH without additional degradation experiments, using a hybrid of domain adaptation and deep neural networks. Another study by Zhu et al. (Zhu et al., 2022) utilized voltage relaxation features to estimate battery capacity through machine learning models, eliminating the need for additional cycling information. Despite their effectiveness, data-driven methods often require extensive labeled data for training, which can be time-consuming and costly to obtain. Additionally, these methods may lack interpretability and generalizability to new battery types or operating conditions.

To address the limitations of purely model-based and data-driven methods, hybrid approaches have been proposed to integrate the strengths of both paradigms. For instance, Thelen et al.(Thelen et al., 2022) presented a lightweight method combining physics-based modeling and machine learning to estimate the capacity and degradation modes of lithium-ion batteries, leveraging both early-life experimental data and simulation data generated from a physics-based half-cell model. Another example is the work by Wang et al. (F. Wang et al., 2024), which proposed a PINN-based approach to model battery degradation dynamics using empirical degradation equations and neural networks. These hybrid methods offer enhanced accuracy and robustness by incorporating physical insights into data-driven models.

Despite the existence of physics-based methods, data-driven approaches, and hybrid methods that integrate both paradigms, these techniques have predominantly focused on learning within a single task. This limitation restricts their ability to fully leverage the interconnections between related tasks. Therefore, Multi-task Learning (MTL) has also been explored to improve the accuracy of SOH and RUL estimation by simultaneously learning related tasks. Kim et al. (Kim & Sohn, 2021) proposed a convolutional neural network-based multitask learning method to integrate health status detection and RUL prediction, demonstrating superior performance on the C-MAPSS dataset compared to traditional models. Similarly, Zhang et al.(S. Zhang et al., 2024) proposed a physics-informed hybrid multitask learning method to simultaneously estimate Li+ concentration dynamics in solid particles and electrolyte, electrode aging states, and battery health status, achieving higher accuracy and interpretability in lithium-ion battery aging estimation. These studies highlight the potential of MTL framework in improving the accuracy and generalizability of battery health monitoring systems.

However, these MTL methods all focus on tasks that will be related but still independent, such as electrode aging states and battery health status, which can lead to poor generalization performance for other datasets in either method. Furthermore, they have difficulty in effectively disentangling shared and task-specific features and solving the interference problem between different tasks, resulting in poor knowledge

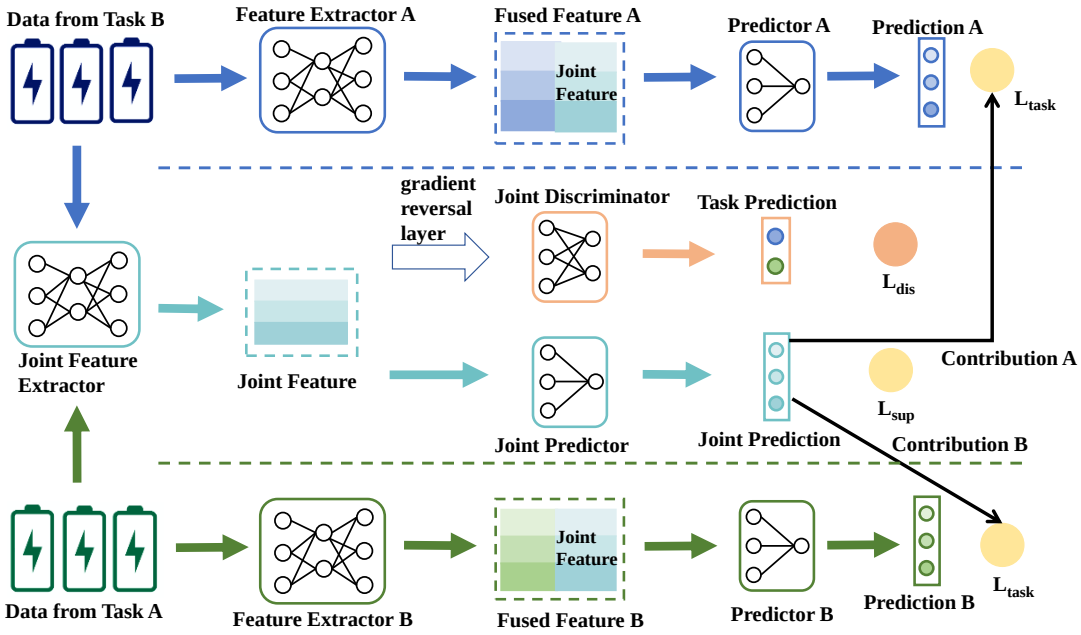


Figure 2. Overview of the proposed SRSE framework for battery SOH estimation.

transfer efficiency. Therefore, we propose a shared representation learning framework for SOH Estimation to address these challenges. The proposed SRSE treats different datasets under the same task as different tasks, to improve the generalization performance of the model and learn multiple datasets at the same time, rather than training a separate network per dataset. The framework includes a joint feature purification mechanism to ensure the shared feature space is free from task-specific contamination. Additionally, we introduce a knowledge distillation module that transfers knowledge from the shared layer to the task-specific layers at both feature and logit levels, thereby enhancing cross-domain transfer efficiency. This dual strategy not only ensures feature purity but also improves model efficiency, offering a robust solution for battery health monitoring tasks.

3. METHODOLOGY

3.1. Problem Formulation

We formulate the cross-configuration SOH estimation via shared representation learning as a tailored multi-task learning problem, where SOH estimation for each battery configuration is treated as a learning task. The overall goal is to train a unified framework that jointly learns multiple tasks for various battery configurations while mitigating task interference and enhancing knowledge transfer (Y. Zhang & Yang, 2021). Formally, let $T = \{T_1, T_2, \ldots, T_M\}$ denote a set of M tasks. Each task T_m is associated with a dataset $\mathcal{D}_m = \{(X_i^m, Y_i^m)\}_{i=1}^{N_m}$ containing N_m samples, where $X_i^m \in \mathbb{R}^d$ is the input vector and Y_i^m is the corresponding label. The objective is to learn a model $f : X \to Y$ that can generalize across all tasks through effectively leveraging shared knowledge(Deng, Chen, Jiang, Song, & Tsang, 2022).

In this work, various battery configurations correspond to different battery types or operating conditions, where each learning task involves predicting the State of Health (SOH) based on recordings of battery characteristics collected during the charge-discharge processes. Specifically, the SOH estimation could be considered as a regression task, whose input X is sensor measurements during a single charge-discharge cycle. The measurements could be structured as multivariate time series sequences with four channels: timestamp, voltage, current, and charge. The output Y of the regression task is a scalar value indicating the SOH, which measures the ratio of the battery's current capacity to its initial capacity.

3.2. Architecture of the Proposed SRSE

Depicted as Fig 2, the proposed Shared Representation learning framework for SOH Estimation (SRSE) consists of a shared layer and M task-specific layers. Each ask-specific layer corresponds to a learning task under one specific battery configuration. Note that we only illustrate the case of two tasks in the diagram, while the framework can exactly handle multiple tasks. The shared layer is trained using data from all tasks simultaneously, allowing it to capture shared representations and knowledge that can be useful across different tasks. On the other hand, the task-specific layers are designed to learn specialized features of each task and output the final label Y(Liu, Qiu, & Huang, 2016). To enhance the overall performance, the proposed framework enables the transfer of both feature-level and logit-level knowledge from the shared layer to the task-specific layers. We introduce a

joint discriminator with an adversarial learning scheme and a joint classifier with a supervised learning scheme to efficiently transfer these two types of shared knowledge.

3.3. Learning the Shared Features with Two Schemes

To effectively learn shared representations across multiple tasks, we introduce two complementary learning schemes: an adversarial learning scheme and a supervised learning scheme. These schemes work together to extract clean and informative joint features while mitigating task interference and noise.

3.3.1. Adversarial Learning for Extracting Pure Joint Features

For the proposed shared representation learning framework, extracting pure joint features is crucial because they contain common, task-invariant information that benefits all tasks while minimizing task-specific biases. However, in practice, the shared feature space often retains task-specific details, leading to interference and suboptimal performance (Liu, Qiu, & Huang, 2017).

In order to tackle this problem, an adversarial learning scheme (Goodfellow et al., 2014) is adopted in SRSE. We first incorporate a Joint Discriminator D, which is a multi-class classifier to identify the task origin of the shared features, i.e., predicts which task a given input feature comes from. Formally, let $f_{\text{shared}}(X)$ denote the shared feature representation extracted from input X. The Joint Discriminator D takes these features as input and predicts the task index m, where $m \in \{1, 2, \ldots, M\}$. The classification loss for D is given by:

$$L_{\text{dis}} = -\mathbb{E}_{(X^m, m)\sim\mathcal{D}} \sum_{m=1}^{M} 1[m] \log D(f_{\text{shared}}(X^m)) \quad (1)$$

where $1[m]$ is the indicator function defined as:

$$1[m] = \begin{cases} 1, & \text{if the task index is } m \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

By minimizing L_{dis}, D learns to correctly classify input features according to their task origin. If these features contain task-specific information, D can easily distinguish them. On the other hand, if the features are truly task-invariant, D will struggle to classify them correctly, indicating that they do not retain task-specific characteristics.

To actively enforce this invariance, we introduce a gradient reversal layer (GRL) between f_{shared} and D(Ganin & Lempitsky, 2015). The GRL negates the gradient from D before propagating it back to f_{shared}, encouraging f_{shared} to learn representations that hinder D's classification ability. Specifically, given an input feature X, the GRL acts as an identity function in the forward pass:

$$\text{GRL}(x) = x, \quad (3)$$

While in the backward pass, it scales the gradient by a negative factor, reversing its direction:

$$\frac{\partial \text{GRL}(x)}{\partial x} = -\alpha I. \quad (4)$$

where I is the identity matrix, and α is a hyperparameter that controls the strength of the gradient reversal. By gradually increasing α, the model initially focuses more on preserving supervised performance, while progressively shifting attention towards minimizing task discrepancy.

This adversarial training enables the simultaneous optimization of f_{shared} and D using Eq. (1), ensuring that the shared representation captures only common and task-invariant information.

3.3.2. Supervised Learning for Preserving Task-Relevant Information

While adversarial learning removes task-specific biases, it also risks discarding useful task-relevant information, leading to performance degradation. To counteract this, we employ a supervised learning scheme that ensures the shared representation remains informative for all tasks. In particular, we introduce a joint predictor P, which is trained to predict a generic label Y based on the shared representation:

$$L_{\text{sup}} = \mathbb{E}_{(X,Y)\sim\mathcal{D}} \left[\ell\left(P(f_{\text{shared}}(X)), Y\right)\right] \quad (5)$$

where $\ell(\cdot)$ is the supervised loss function, such as mean squared error (MSE) for regression tasks or cross-entropy loss for classification tasks. This supervision ensures that the extracted shared features retain task-relevant knowledge rather than collapsing into representations dominated by noise.

3.4. Enhancing Task-Specific Learning with Shared Knowledge

To improve the robustness and generalizability of task-specific layers, we develop two knowledge enhancement strategies: feature-level fusion and logit-level augmentation. These two approaches both optimize the task-specific layers, ensuring that they benefit from the shared representation, while still leveraging the inherent, task-specific information(Xu, Wu, Li, Mao, & Chen, 2023).

3.4.1. Feature-Level Fusion

Feature-level fusion aims to combine shared and task-specific features to produce more informative representations for each task. In our approach, we use element-wise addition as the fusion technique. Mathematically, let $f_{\text{task}}(X)$ be the task-specific features. The fused feature vector F_{fused} is computed

as:

$$F_{\text{fused}} = f_{\text{shared}}(X) + f_{\text{task}}(X) \tag{6}$$

This allows both sources of information to be integrated straightforwardly and efficiently. By utilizing feature-level fusion, we enable the model to generalize better across tasks while simultaneously capturing the unique aspects of each learning task, thereby enhancing the model's overall performance.

3.4.2. Logit-Level Augmentation

While feature-level fusion integrates knowledge at the representation stage, logit-level augmentation refines task-specific predictions by incorporating additional logit-level guidance from the shared layer into the learning process. Specifically, it leverages the predictive confidence of the shared representation to guide task-specific learning (Hinton, Vinyals, & Dean, 2015). During the supervised training of the shared layer, we can estimate how much the shared features contribute to each sample's prediction. This contribution is derived from the joint prediction $P(f_{\text{shared}}(X))$, which, in classification tasks, corresponds to the probability distribution over true labels:

$$w_i = P(Y_i \mid f_{\text{shared}}(X_i)) \tag{7}$$

and in regression tasks, represents the error relative to the ground truth:

$$w_i = \exp\left(-\frac{|P(f_{\text{shared}}(X_i)) - Y_i|}{\beta}\right) \tag{8}$$

where β is a scaling factor that controls the sensitivity of the weight adjustment.

To incorporate this knowledge into task-specific learning, we assign w_i for each sample i to control its influence on the task-specific layer's training:

$$L_{\text{task}} = \frac{1}{N}\sum_{i=1}^{N} w_i \ell(P_{\text{task}}(F_{\text{fused}}), Y_i) \tag{9}$$

where $\ell(\cdot)$ is the supervised loss function, and P_{task} represents the task-specific predictor. This dynamic weighting mechanism ensures that task-specific layers pay more attention to samples that provide more reliable task-invariant information, thereby preventing the model from overfitting to task-specific features.

Finally, the complete details of our proposed SRSE framework are illustrated in Algorithm 1.

4. EXPERIMENTS

4.1. Datasets

To evaluate the performance of our proposed framework, we conduct experiments on three different large-scale datasets in the field of battery health, including the TJU dataset (Zhu et al., 2022), the XJTU dataset (F. Wang et al., 2024), and the SANDIA dataset (Preger et al., 2020). These datasets cover a wide range of lithium-ion battery types, chemistries, and operating conditions, with the aim of predicting the remaining life of a battery based on the curves of voltage, current, and charge over time during charging and discharging.

Each dataset contains detailed charge/discharge profiles and corresponding state of health (SOH) information, enabling us to thoroughly evaluate the accuracy and stability of our model under different scenarios. In each dataset, we divide the tasks based on the chemical composition of the battery, charging protocol, and other factors. Below is a detailed description of each dataset.

4.1.1. TJU

The TJU dataset contains three types of batteries: NCA battery (3500 mAh nominal capacity and 2.65–4.2 V cut-off voltage), NCM battery (3500 mAh nominal capacity and 2.5–4.2 V cut-off voltage), and NCM + NCA battery (2500 mAh nominal capacity and 2.5–4.2 V cut-off voltage). The batteries were cycled at different temperatures (25°C, 35°C, and 45°C) and various charge/discharge rates (0.25C to 4C). We use batch 1, batch 2, and batch 3 to represent NCA, NCM, and NCM + NCA batteries, respectively.

4.1.2. XJTU

The XJTU battery dataset consists of 55 Lithium Nickel Cobalt Manganese Oxide (NCM) batteries with a nominal capacity of 2000 mAh, a nominal voltage of 3.6 V, and the cut-off voltages for charging and discharging are 4.2 V and 2.5 V. The batteries were cycled to failure under six distinct charge/discharge protocols at room temperature. These protocols include fixed charging and discharging, random discharging with a fixed current in different cycles, random walking, and the charging and discharging strategy of a satellite in Geosynchronous Earth Orbit (GEO). In our experiments, we use batch 1 to batch 4 to respectively represent the first 4 charging/discharging protocols in (F. Wang et al., 2024).

4.1.3. SANDIA

The Sandia dataset is derived from a comprehensive study conducted by Sandia National Laboratories, examining the degradation of commercial 18650 lithium-ion cells with different chemistries (NCA, NMC, and LFP) under various cycling conditions, including different temperatures, depths of discharge (DOD), and discharge rates, to investigate their long-term degradation behavior. In our experiments, we use batches 1 to 3 to represent LFP-15C, LFP-25C, and LFP-35C batteries, respectively.

Algorithm 1 The Proposed SRSE Framework

Input: Multi-task dataset $\mathcal{D} = \{\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_M\}$, learning rates η_1, η_2, gradient reversal factor α, weight scaling factor β.
Output: Optimized task-specific extractors f^m_{task} and predictors P^m_{task} for each task T_m.

Initialize shared feature extractor f_{shared}, task-specific extractors f^m_{task}, task predictors P^m_{task}, joint predictor P, and joint discriminator D.
repeat
  for each mini-batch $\{(X^m_i, Y^m_i)\}_{i=1}^B$ from all tasks T_m **do**
    // Step 1: Extract shared features
    $F_{\text{shared}} \leftarrow f_{\text{shared}}(X)$
    // Step 2: Adversarial training with Joint Discriminator
    Compute discriminator loss: $L_{\text{dis}} \leftarrow -\frac{1}{M}\sum_{m=1}^{M} 1[m] \log D(F_{\text{shared}})$
    Update discriminator D: $D \leftarrow D - \eta_1 \nabla_D L_{\text{dis}}$
    Reverse gradients via GRL and update f_{shared}: $f_{\text{shared}} \leftarrow f_{\text{shared}} + \eta_1 \alpha \nabla_{f_{\text{shared}}} L_{\text{dis}}$
    // Step 3: Supervised learning for shared feature
    Compute supervised loss: $L_{\text{sup}} \leftarrow \frac{1}{B}\sum_{i=1}^{B} \ell(P(F_{\text{shared}}), Y_i)$
    Update shared feature extractor f_{shared}: $f_{\text{shared}} \leftarrow f_{\text{shared}} - \eta_1 \nabla_{f_{\text{shared}}} L_{\text{sup}}$
    // Step 4: Task-specific learning with feature-level fusion
    Extract task-specific features: $F^m_{\text{task}} \leftarrow f^m_{\text{task}}(X)$
    Fuse features: $F^m_{\text{fused}} \leftarrow F_{\text{shared}} + F^m_{\text{task}}$
    // Step 5: Task-specific learning with logit-level augmentation
    Compute sample weights: $w_i \leftarrow \exp(-|P(F_{\text{shared},i}) - Y_i|/\beta)$
    Compute task-specific loss: $L^m_{\text{task}} \leftarrow \frac{1}{B}\sum_{i=1}^{B} w_i \ell(P^m_{\text{task}}(F^m_{\text{fused}}), Y_i)$
    Update task-specific predictor: $P^m_{\text{task}} \leftarrow P^m_{\text{task}} - \eta_2 \nabla_{P^m_{\text{task}}} L^m_{\text{task}}$
    Update task-specific extractor: $f^m_{\text{task}} \leftarrow f^m_{\text{task}} - \eta_2 \nabla_{f^m_{\text{task}}} L^m_{\text{task}}$
  end for
until convergence or maximum iterations reached

4.2. Data Preprocessing

To ensure the consistency and reliability of the dataset, we implemented a rigorous data preprocessing pipeline. The preprocessing steps include extracting key cycle-level features, filtering outliers, aligning timestamps, and normalizing data for model input.

4.2.1. Data Extraction and Cleaning

We first process the raw dataset by grouping data based on the cycle number. Each cycle's data is extracted separately, and the corresponding state of health (SOH) is calculated as:

$$SOH = \frac{Q_{\text{discharge}}}{Q_{\text{nominal}}} \tag{10}$$

where $Q_{\text{discharge}}$ is the maximum discharge capacity observed in the cycle, and Q_{nominal} is the nominal capacity of the battery. Cycles with SOH values outside the range of [0.5, 2] are considered outliers and removed from further analysis.

4.2.2. Feature Extraction and Temporal Interpolation

To characterize battery behavior during charging, we extract key features from the constant-current constant-voltage (CCCV) phase, including voltage, current, charge capacity, and timestamp. Given that the length of CCCV data varies across cycles due to differences in sampling rates and operating conditions, we standardize each cycle's data using linear interpolation.

To ensure uniformity, we resample each cycle's time series to a fixed length of 1000 points by evenly dividing the total duration into 1000 intervals:

$$t_{\text{interp}} = t_{\min} + k \cdot \frac{t_{\max} - t_{\min}}{999}, \quad k = 0, 1, \ldots, 999 \tag{11}$$

$$V_{\text{interp}} = \text{interp}(t_{\text{interp}}, t, V) \tag{12}$$

$$I_{\text{interp}} = \text{interp}(t_{\text{interp}}, t, I) \tag{13}$$

$$Q_{\text{interp}} = \text{interp}(t_{\text{interp}}, t, Q) \tag{14}$$

where $\text{interp}(\cdot)$ denotes linear interpolation. Additionally, we compute the charge capacity differential concerning voltage:

$$\frac{dQ}{dV} = \frac{\Delta Q_{\text{interp}}}{\Delta V_{\text{interp}}} \tag{15}$$

After preprocessing, each charge-discharge cycle is represented as a multivariate time series with four channels: volt-

age, current, charge capacity, and dQ/dV, all standardized to 1000 time steps.

4.2.3. Final Dataset Structure

Each cycle is represented as a time-series feature matrix of shape $[5, 1000]$, where the five channels correspond to interpolated voltage, current, charge capacity, charge derivative (dQ/dV), and time. The final dataset is stored as PyTorch tensors for efficient model training.

4.3. Experiments Setup

For the proposed SRSE framework, we select different batches from the dataset to represent different tasks. For example, in the TJU dataset, batches corresponding to different chemical components and different charge/discharge protocols are treated as separate tasks. In addition, both the joint network and the task-specific networks employ different models as their backbone, including 1-D CNN, MLP, DNN, and LSTM.

During training, a well-trained joint network is a prerequisite to perform task-invariant information distillation. We first employ the domain-adversarial training of neural networks (DANN) approach (Ganin et al., 2016) to train the joint network. Once the joint network is trained, we freeze its parameters to prevent further updates. We then train the task-specific networks on top of the frozen joint network. This allows each task-specific network to focus on learning task-specific patterns while leveraging the joint feature representations provided by the joint network. Furthermore, in terms of evaluation metrics, we adopt MAPE and RMSE for all experiments. For all experiments, we repeat 10 times with different random seeds.

4.4. Main Experiments Results and Comparative Analysis

We first employ a 1-D CNN with nine layers as the backbone on three datasets. Table 1 shows that using the same CNN and the same hyperparameter configuration, our proposed SRSE framework achieves better performance than single-task training in the vast majority of tasks, with a more significant lead. Performance improvements can reach up to 69.05%. Across all tasks with decreased errors, RMSE metrics decreased by an average of 46.71% and MAPE metrics decreased by an average of 46.58%. Furthermore, even in tasks where the minimum error was not achieved, our method exhibits only a minimal gap compared to the best results obtained.

In addition, to verify the effectiveness of our proposed SRSE framework based on different backbones, we replace different backbones on the TJU dataset, including MLP, DNN, and LSTM. From Table 3, it can be seen that our SRSE framework still works despite changing different backbones. Our SRSE framework always achieves the best performance among all backbone approaches. This significantly demonstrates the effectiveness of our proposed SRSE framework in enhancing the performance of individual tasks. The consistent improvements across various metrics highlight the robustness and versatility of our approach, which is capable of elevating the performance even in scenarios where single-task training has struggled to achieve optimal results. This underscores the potential of our SRSE framework as a valuable tool for optimizing task-specific outcomes within a shared representation learning context.

4.5. SRSE Across Different Operating Conditions

In 4.4, we evaluated our SRSE framework by partitioning tasks based on intrinsic battery properties, such as material composition and model type. In this section, we further investigate the effectiveness of our framework by defining tasks based on external operating conditions, including charging rate, discharging rate, and temperature. This setting allows us to assess how well the framework can leverage shared knowledge across different environmental conditions, which is crucial for real-world battery health monitoring applications.

To construct these tasks, we focus on the TJU dataset and partition it based on specific external conditions:

- **Charging Rate**: We select cells from the NCA_25°C batch and categorize them into three tasks based on different charge rates (1C, 0.5C, 0.25C), keeping all other conditions constant.
- **Discharging Rate**: We select cells from the NCM and NCA batches and define three tasks corresponding to different discharge rates (1C, 2C, 4C).
- **Temperature**: We select cells from the NCM batch and divide them into three tasks based on ambient temperatures (25°C, 35°C, 45°C), keeping all other factors such as charge-discharge rates and battery chemistry unchanged.

Table 2 presents the performance comparison between the SRSE (with CNN as backbone) model and the single-task CNN model across these different external conditions. The results demonstrate that SRSE consistently outperforms the single-task model in most cases, achieving lower RMSE and MAPE. Notably, in the charge rate setting, SRSE consistently outperforms the single-task model across all tasks. Furthermore, for tasks defined by discharge rate variations, our model exhibits a substantial performance gain, particularly at 2C and 4C, where RMSE is reduced by 61.6% and 16.2%, respectively. This suggests that SRSE effectively captures common patterns across different environmental conditions, enabling more accurate and robust SOH estimations.

Table 1. The results of our proposed SRSE framework on three datasets. The best results are in bold.

Dataset	Task	SRSE		CNN		Improvement	
		RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
TJU	1	**0.0155**	**0.0153**	0.0201	0.0196	22.89%	21.94%
	2	**0.0137**	**0.0141**	0.0200	0.0208	31.5%	32.21%
	3	**0.0176**	**0.0167**	0.0357	0.0352	50.7%	52.56%
XJTU	1	**0.0191**	**0.0173**	0.0558	0.0559	65.77%	69.05%
	2	**0.0257**	**0.0247**	0.0586	0.0590	56.14%	58.14%
	3	**0.0237**	**0.0221**	0.0625	0.0630	62.08%	64.92%
	4	**0.0187**	**0.0155**	0.0514	0.0508	65.37%	69.49%
SANDIA	1	**0.0169**	0.0166	0.0270	**0.0153**	37.41%	-
	2	0.0086	**0.0052**	**0.0061**	**0.0052**	-	-
	3	**0.0180**	**0.0176**	0.0252	0.0184	28.57%	4.35%

Table 2. Comparisons of SRSE and CNN across different operating conditions.

Dataset	condition	task	meaning	SRSE		CNN	
				RMSE	MAPE	RMSE	MAPE
NCA_25°C	charge rate	1	1C	**0.0130**	**0.0137**	0.0146	0.0143
		2	0.5C	**0.0096**	**0.0088**	0.0110	0.0114
		3	0.25C	**0.0109**	**0.0108**	0.0193	0.0199
NCM/NCA	discharge rate	1	1C	0.0176	**0.0077**	**0.0175**	0.0098
		2	2C	**0.0118**	**0.0090**	0.0307	0.0304
		3	4C	**0.0129**	**0.0083**	0.0154	0.0109
NCM	temperature	1	25°C	**0.0074**	**0.0088**	0.0117	0.0126
		2	35°C	**0.0061**	**0.0054**	0.0075	0.0073
		3	45°C	0.0099	0.0112	**0.0097**	**0.0107**

However, we also observe that in some cases, such as high-temperature conditions (45°C), the performance gap between shared representation learning across multiple battery configurations and configuration-specific learning is minimal. This may be attributed to the accelerated aging effects at high temperatures, which introduce complex degradation patterns and make knowledge transfer less effective.

Overall, these results confirm that our proposed SRSE framework enhances predictive accuracy under varying external conditions, thereby enabling more reliable SOH estimation and improving battery health management in diverse operational scenarios.

4.6. Ablation Study

There are two key components in our proposed framework: the joint discriminator D and feature enhancement of the task-specific layer(contribution). To analyze the contribution of each component, we conduct an ablation study on the TJU dataset using CNN as the backbone. As Table 4 shows, first, without using the joint discriminator and the feature enhancement of the task-specific layer, there are only shared features passed to each task-specific layer. As can be seen from the results, some tasks may have relatively poor performance due to noise in the shared features or interference from other tasks. Then, we add adversarial training, or a joint discriminator, so that the shared features contain task-invariant information, and the shared features are more "pure", thus achieving better performance. Finally, by adding all the components, our approach achieves optimal performance on all three tasks. This shows that all our proposed components work and contribute to the performance of each task. This demonstrates the efficiency of the proposed SRSE framework, with each core component contributing positively to the overall performance enhancement. The joint discriminator aids in ensuring the shared features are more refined and contain task-invariant information, while the feature enhancement mechanism further optimizes these features, leading to superior model performance across various tasks.

5. CONCLUSION

In this work, we propose a novel shared representation learning framework for generalizable SOH Estimation across multiple Battery configurations, which effectively addresses the limitations of existing methods by leveraging the inherent correlations between different battery types. Through an adversarial training scheme, our approach mitigates task interference in the shared feature space, ensuring better generalization across diverse battery types and operating conditions. Additionally, transferring both feature-level and logit-level knowledge to task-specific layers further enhances the adaptability and efficiency of the model. Experimental results on three large-scale battery health datasets confirm that our method significantly improves generalization and predictive capability, offering a more robust and accurate solution for battery health monitoring in real-world applications.

REFERENCES

Attia, P. M., Grover, A., Jin, N., Severson, K. A., Markov, T. M., Liao, Y.-H., ... others (2020). Closed-

Table 3. Comparison between our proposed framework of different backbones on the TJU Dataset. The best results are in bold.

Dataset	Task	**MLP**		**SRSE**		**DNN**		**SRSE**		**LSTM**		**SRSE**	
		RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
TJU	1	0.0246	0.0229	**0.0107**	**0.0109**	0.0246	0.0229	**0.011**	**0.0083**	0.0246	0.0229	**0.0171**	**0.0153**
	2	0.0186	0.0192	**0.0121**	**0.0115**	0.0186	0.0192	**0.0123**	**0.0121**	0.0186	0.0192	**0.0115**	**0.0098**
	3	0.0395	0.0401	**0.0143**	**0.0081**	0.0395	0.0401	**0.0148**	**0.013**	0.0395	0.0401	**0.0224**	**0.012**

Table 4. Ablation study on the TJU dataset. The best results are in bold.

Dataset	Task	**proposed w/o con+dis**		**proposed w/o con**		**proposed**	
		RMSE	**MAPE**	**RMSE**	**MAPE**	**RMSE**	**MAPE**
TJU	1	0.0179	0.0187	0.0156	0.0155	**0.0155**	**0.0153**
	2	0.0147	0.0155	0.0156	0.0162	**0.0137**	**0.0141**
	3	0.0337	0.0361	0.0209	0.0205	**0.0176**	**0.0167**

loop optimization of fast-charging protocols for batteries with machine learning. *Nature*, *578*(7795), 397–402. Retrieved from https://www.nature.com/articles/s41586-020-1994-5

Baghdadi, I., Briat, O., Delétage, J.-Y., Gyan, P., & Vinassa, J.-M. (2016). Lithium battery aging model based on dakin's degradation approach. *Journal of Power Sources*, *325*, 273–285. Retrieved from https://www.sciencedirect.com/science/article/abs/pii/S0378775316307388

Ben-David, S., Blitzer, J., Crammer, K., Kulesza, A., Pereira, F., & Vaughan, J. W. (2010). A theory of learning from different domains. *Machine learning*, *79*, 151–175. Retrieved from https://link.springer.com/article/10.1007/s10994-009-5152-4

Ben-David, S., Blitzer, J., Crammer, K., & Pereira, F. (2006). Analysis of representations for domain adaptation. *Advances in neural information processing systems*, *19*. Retrieved from https://proceedings.neurips.cc/paper/2006/hash/b1b0432ceafb0ce714426e9114852ac7-Abstract.html

Chen, S., Zhang, Y., & Yang, Q. (2024). Multi-task learning in natural language processing: An overview. *ACM Computing Surveys*, *56*(12), 1–32. Retrieved from https://dl.acm.org/doi/abs/10.1145/3663363

Deng, J., Chen, X., Jiang, R., Song, X., & Tsang, I. W. (2022). A multi-view multi-task learning framework for multi-variate time series forecasting. *IEEE Transactions on Knowledge and Data Engineering*, *35*(8), 7665–7680. Retrieved from https://ieeexplore.ieee.org/abstract/document/9935292

Dong, G., & Wei, J. (2021). A physics-based aging model for lithium-ion battery with coupled chemical/mechanical degradation mechanisms. *Electrochimica Acta*, *395*, 139133. Retrieved from https://www.sciencedirect.com/science/article/abs/pii/S0013468621014237

Ganin, Y., & Lempitsky, V. (2015). Unsupervised domain adaptation by backpropagation. In *International conference on machine learning* (pp. 1180–1189). Retrieved from https://proceedings.mlr.press/v37/ganin15

Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., ... Lempitsky, V. (2016). Domain-adversarial training of neural networks. *Journal of machine learning research*, *17*(59), 1–35. Retrieved from https://www.jmlr.org/papers/v17/15-239.html

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... Bengio, Y. (2014). Generative adversarial nets. *Advances in neural information processing systems*, *27*. Retrieved from https://proceedings.neurips.cc/paper_files/paper/2014/hash/f033ed80deb0234979a61f95710dbe25-Abstract.html

He, L., Li, H., Wang, J., Chen, M., Gozdas, E., Dillman, J. R., & Parikh, N. A. (2020). A multitask, multi-stage deep transfer learning model for early prediction of neurodevelopment in very preterm infants. *Scientific Reports*, *10*(1), 15072. Retrieved from https://www.nature.com/articles/s41598-020-71914-x

Heuer, F., Mantowsky, S., Bukhari, S., & Schneider, G. (2021). Multitask-centernet (mcn): Efficient and diverse multitask learning using an anchor free approach. In *Proceedings of the ieee/cvf international conference on computer vision* (pp. 997–1005). Retrieved from https://openaccess.thecvf.com/content/ICCV2021W/ERCVAD/html/Heuer_MultiTask-CenterNet_MCN_Efficient_and_Diverse_Multitask_Learning_Using_an_Anchor_ICCVW_2021_paper.html

Hinton, G., Vinyals, O., & Dean, J. (2015). Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*. Retrieved from https://arxiv.org/abs/1503.02531

Hou, Y., Ragab, M., Wang, Y., Wu, M., Alseiari, A., Kwoh, C.-K., ... Chen, Z. (2025). Evidential domain adaptation for remaining useful life prediction with incomplete degradation. *IEEE Transactions on Instrumentation and Measurement*. Retrieved from https://ieeexplore.ieee.org/abstract/document/10930593

Hou, Y., Wang, Y., Wu, M., Kwoh, C.-K., Li, X., & Chen, Z. (2025). Source-free calibrated uncertainty for rul adaptation with incomplete degradation. *Mechanical Systems and Signal Processing*, *233*, 112718. Retrieved from https://www.sciencedirect.com/science/article/abs/pii/S0888327025004194

Kim, T. S., & Sohn, S. Y. (2021). Multitask learning for health condition identification and remaining useful life prediction: Deep convolutional neural network approach. *Journal of Intelligent Manufacturing*, *32*(8), 2169–2179. Retrieved from https://link.springer.com/article/10.1007/s10845-020-01630-w

Li, X., Lyv, M., Gao, X., Li, K., & Zhu, Y. (2025). A co-estimation framework of state of health and remaining useful life for lithium-ion batteries using the semi-supervised learning algorithm. *Energy and AI*, *19*, 100458. Retrieved from https://www.sciencedirect.com/science/article/pii/S2666546824001241

Liu, P., Qiu, X., & Huang, X. (2016). Recurrent neural network for text classification with multi-task learning. *arXiv preprint arXiv:1605.05101*. Retrieved from https://arxiv.org/abs/1605.05101

Liu, P., Qiu, X., & Huang, X. (2017). Adversarial multi-task learning for text classification. *arXiv preprint arXiv:1704.05742*. Retrieved from https://arxiv.org/abs/1704.05742

Lu, J., Xiong, R., Tian, J., Wang, C., & Sun, F. (2023). Deep learning to estimate lithium-ion battery state of health without additional degradation experiments. *Nature Communications*, *14*(1), 2760. Retrieved from https://www.nature.com/articles/s41467-023-38458-w

Lui, Y. H., Li, M., Downey, A., Shen, S., Nemani, V. P., Ye, H., ... others (2021). Physics-based prognostics of implantable-grade lithium-ion battery for remaining useful life prediction. *Journal of Power Sources*, *485*, 229327. Retrieved from https://www.sciencedirect.com/science/article/abs/pii/S0378775320316153

Preger, Y., Barkholtz, H. M., Fresquez, A., Campbell, D. L., Juba, B. W., Romàn-Kustas, J., ... Chalamala, B. (2020). Degradation of commercial lithium-ion cells as a function of chemistry and cycling conditions. *Journal of The Electrochemical Society*, *167*(12), 120532. Retrieved from https://iopscience.iop.org/article/10.1149/1945-7111/abae37

Rauf, H., Khalid, M., & Arshad, N. (2022). Machine learning in state of health and remaining useful life estimation: Theoretical and technological development in battery degradation modelling. *Renewable and Sustainable Energy Reviews*, *156*, 111903. Retrieved from https://www.sciencedirect.com/science/article/abs/pii/S1364032121011692

Thelen, A., Lui, Y. H., Shen, S., Laflamme, S., Hu, S., Ye, H., & Hu, C. (2022). Integrating physics-based modeling and machine learning for degradation diagnostics of lithium-ion batteries. *Energy Storage Materials*, *50*, 668–695. Retrieved from https://www.sciencedirect.com/science/article/abs/pii/S2405829722002999

Wang, F., Zhai, Z., Zhao, Z., Di, Y., & Chen, X. (2024). Physics-informed neural network for lithium-ion battery degradation stable modeling and prognosis. *Nature Communications*, *15*(1), 4332. Retrieved from https://www.nature.com/articles/s41467-024-48779-z

Wang, Y., Wu, M., Li, X., Xie, L., & Chen, Z. (2025). A survey on graph neural networks for remaining useful life prediction: Methodologies, evaluation and future trends. *Mechanical Systems and Signal Processing*, *229*, 112449. Retrieved from https://www.sciencedirect.com/science/article/abs/pii/S0888327025001505

Wen, Z., Fang, Y., Wei, P., Liu, F., Chen, Z., & Wu, M. (2025). Temporal and heterogeneous graph neural network for remaining useful life prediction. *IEEE Transactions on Neural Networks and Learning Systems*. Retrieved from https://ieeexplore.ieee.org/abstract/document/11106703

Xu, Q., Wu, M., Li, X., Mao, K., & Chen, Z. (2023). Distilling universal and joint knowledge for cross-domain model compression on time series data. *arXiv preprint arXiv:2307.03347*. Retrieved from https://arxiv.org/abs/2307.03347

Zhang, S., Liu, Z., Xu, Y., & Su, H. (2024). A physics-informed hybrid multitask learning for lithium-ion battery full-life aging estimation at early lifetime. *IEEE Transactions on Industrial Informatics*. Retrieved from https://ieeexplore.ieee.org/abstract/document/10691675

Zhang, Y., & Yang, Q. (2021). A survey on multi-task learning. *IEEE transactions on knowledge and data engineering*, *34*(12), 5586–5609. Retrieved from https://ieeexplore.ieee.org/abstract/document/9392366

Zhu, J., Wang, Y., Huang, Y., Bhushan Gopaluni, R., Cao, Y., Heere, M., ... others (2022). Data-driven capac-

ity estimation of commercial lithium-ion batteries from voltage relaxation. *Nature communications*, *13*(1), 2261. Retrieved from https://www.nature.com/articles/s41467-022-29837-w