

早期循环数据对锂离子电池质量分选的MAE预测方法研究

陆天鹏¹, 沈逸凡¹, 吕桃林², 周 龙¹

¹上海理工大学机械工程学院, 上海

²上海空间电源研究所, 空间电源全国重点实验室, 上海

收稿日期: 2025年4月27日; 录用日期: 2025年5月20日; 发布日期: 2025年5月29日

摘 要

锂离子电池的早期质量分类对提升电池整体质量至关重要。然而, 仅依赖于预测电池单体的寿命终止(EOL)并不足够。由于锂离子电池老化机理多样、化学变化复杂, 以及环境和使用条件的差异, 导致相同使用条件下电池单体的老化速率仍然存在显著差异。这一问题直接影响了电池早期质量分类的准确性, 并降低了后续电池成组的安全性和可靠性。为应对这一挑战, 研究生成了一个数据集, 包括154个在相同加速老化条件下循环的18650三元锂离子电池。并提出了一种基于半监督学习的寿命终止(EOL)预测方法。首先利用大量无标签的前3个循环数据对生成式无监督结构掩码自编码器(MAE)进行预训练, 以实现自动特征提取, 避免了充放电数据人工特征提取的失真和高成本。随后, 通过少量标记数据微调预训练模型, 最终实现高精度EOL预测。评估结果表明, 该方法具有高度准确性: 仅使用20%的标记数据即可将验证集上的RMSE降低至2.3%。该研究为利用未标记数据进行电池早期质量分类提供了一种新方法, 展示了其应用前景。

关键词

锂离子电池, 掩码自编码器, 半监督学习, 寿命中止

Research on MAE Prediction Method for Early Cycle Data-Based Quality Sorting of Lithium-Ion Batteries

Tianpeng Lu¹, Yifan Shen¹, Taolin Lyu², Long Zhou¹

¹School of Mechanical Engineering, University of Shanghai for Science and Technology, Shanghai

²State Key Laboratory of Space Power-Sources, Shanghai Institute of Space Power-Sources, Shanghai

Received: Apr. 27th, 2025; accepted: May 20th, 2025; published: May 29th, 2025

文章引用: 陆天鹏, 沈逸凡, 吕桃林, 周龙. 早期循环数据对锂离子电池质量分选的 MAE 预测方法研究[J]. 建模与仿真, 2025, 14(5): 855-867. DOI: 10.12677/mos.2025.145439

Abstract

Early quality classification of lithium-ion batteries is crucial for improving their overall performance. However, relying solely on end-of-life (EOL) prediction of individual cells is insufficient. Due to the diverse aging mechanisms, complex chemical changes, and variations in environmental and usage conditions, even cells operated under identical accelerated aging protocols can exhibit significantly different degradation rates. This variation directly affects the accuracy of early quality classification and compromises the safety and reliability of battery pack assembly. To address this challenge, a dataset comprising 154 18650-type ternary lithium-ion batteries cycled under identical accelerated aging conditions was developed. A semi-supervised learning approach is proposed for EOL prediction. Specifically, a generative Masked Autoencoder (MAE) is pre-trained using abundant unlabeled data from the first three cycles to enable automatic feature extraction, thereby avoiding the distortion and high costs associated with manual feature engineering from charge/discharge data. The pre-trained model is then fine-tuned using a small amount of labeled data to achieve high-accuracy EOL prediction. Evaluation results show that this method is highly effective: using only 20% of the labeled data reduces the RMSE on the validation set to 2.3%. This study presents a novel approach for early quality classification of lithium-ion batteries using unlabeled data and demonstrates promising application potential.

Keywords

Lithium-Ion Battery, Masked Autoencoder, Semi-Supervised Learning, End of Life (EOL)

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

在全球追求绿色和低碳未来的背景下,锂离子电池(LIB)在现代电力交通系统(如新能源电动汽车(EV)、电动火车等)的能量储存和供应中发挥着重要作用[1] [2]。然而,在 LIB 的进一步发展中仍然存在许多关键问题需要解决,例如研究和开发(R&D)的高成本以及安全性问题[3]。研究快速预测电池寿命和通过早期劣化循环对电池质量进行分类的方法,对于加快电池研发周期和保证电池产品质量具有重要意义。

为了满足现代电动汽车的快速充电需求,电池研发中的一个关键研究方向是多步快速充电设计和优化,其目的是确定最佳快速充电曲线,以最大限度地减少电池充电时间,同时最大限度地延长电池寿命[4]。然而,电池生命周期测试是耗时的,这对优化过程形成了重大障碍。例如,寿命为 2000 次循环的电池可能需要几个月才能失效。早期循环中的快速电池寿命预测和质量分类旨在加速电池设计和优化[5]。例如,只需要前 5 个循环数据作为输入的技术可以快速地将测试电池分类为长寿命的好电池或短寿命的坏电池。基于该技术,可以快速验证不同装药剖面的性能,提高优化效率,降低研发成本。此外,在电池制造中,早期循环寿命分类可以实现对异常过短寿命电池的快速识别,另一方面可以确保交付给 EV 市场的电池质量[6]。为了满足电动汽车更高的功率要求,LIB 电池通常串联或并联连接以形成电池组[7]。然而,在电池组中,不可避免地存在电池不一致性,例如不同的标称容量、热分布、退化模式、寿命等。这种不一致性会显著影响整个电池组的性能,因为电池组寿命通常由电池组中寿命最短的电池单元决定[8]。因此,在电池组组装阶段,快速识别具有相似寿命和退化模式的电池单元,并将它们分配到同一组

中,可以有效地缓解电池不一致性。最近,锂离子电池的回收和二次寿命应用也引起了越来越多的关注,因为这种新的应用可以有效地延长电池的使用寿命,更好地利用资源,提高经济效益[9]。最新报告显示[10],到2030年,每年将有超过600万个锂电池组从电动汽车上退役。早期循环电池质量分类技术可以作为此类应用中的重要工具,因为其能够将退役的LIB分类到不同的寿命组中,然后更好地将其适合于不同场景下的二次寿命应用。因此,发展早期循环电池质量分级技术具有重要的价值和广阔的应用前景。然而,在这方面的研究在文献中是不够的。

为解决上述问题,本文提出了一种基于MAE的LIB快速预测全生命周期EOL。该模型由两个主要部分组成:首先是使用大量无标签的前三个循环的电压、电流和SOC数据预训练生成初步预训练模型,此模型初步具备自动特征提取的能力。然后,使用少量带标签的数据对预训练模型进行微调,最终实现准确的全生命周期EOL预测。此过程需要进行一系列对比实验去选择验证。首先,是对输入数据参数量的选择,数据参数量的选择对训练时间及训练精度有着较高的影响。其次,是对超参数规模的选择,超参数规模的选择对于训练时间、设备需求有着较大影响。最后,是对早期循环数据的选择,我们需要尽可能地使用少量数据去进行EOL预测,同时也要保证预测的准确性。

2. 实验介绍及数据预处理

2.1. 锂离子电池基本信息及实验条件

首先对本文在研究中所用电池的基本信息进行介绍,如图1所示,这款电池是江西远东电池有限公司生产的FEB18650-2500mAh的18650圆柱型电池,该电池的基本参数信息列于表1中。



Figure 1. 18650 cylindrical battery
图 1. 18650 圆柱型电池

实验设备如图2所示,其中:图2(a)为BTH-150C高低温交变试验箱;图2(b)为本研究所用的同批次18650圆柱型电池;图2(c)为新威CT-4008-5V12A-DB高精度电池测试系统,用于电池的充放电测试;图2(d)为AutoLab电化学工作站,包括上位机、恒电位仪和放大器;图2(e)为上位机。在实验过程中,需先调节温箱并保持电池实验温度。循环老化实验开始前,电池需放置于温箱中静置,以确保温度均衡。电池性能测试系统通过线束连接电池与上位机,上位机可发送指令控制实验设备,并记录电流、电压、温度等实验数据。所有EIS测试均需在恒温箱内进行。测试前,需将电池与电化学工作站通过线束连接,并在恒温箱内静置不少于3小时,以确保电池内部和外部温度平衡。电化学工作站通过向电池施加电流/

电压激励，并采集电压/电流反馈信号，从而获得阻抗谱数据。

Table 1. Basic performance parameters of the experimental battery
表 1. 实验电池基本性能参数

电池参数名称	数值及类别
正 负极材料	三元锂 石墨(LiNi _x Mn _y Co _z O ₂ C ₆)
额定容量(mAh)	2500
标称电压(V)	3.6
充电截止电压(V)	4.2
放电截止电压(V)	2.75
标准充/放电流(A)	1.25 (0.5 C)/2.5 (1 C)

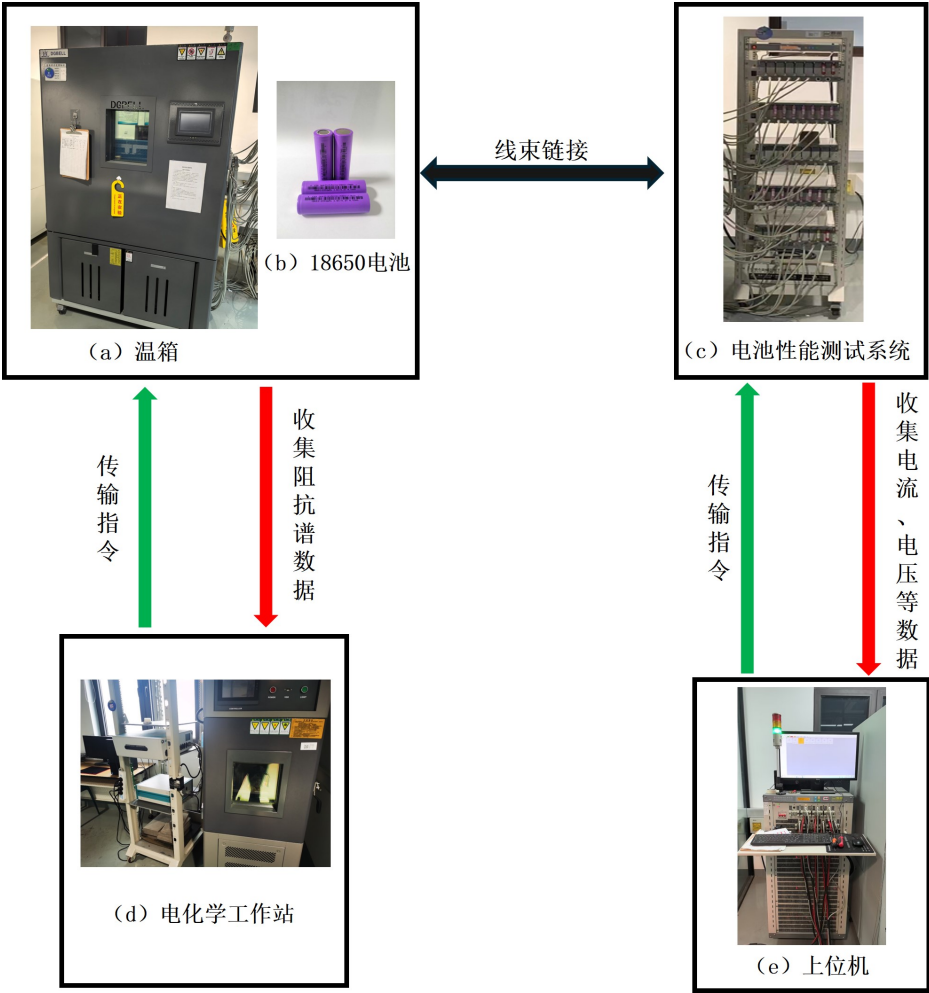


Figure 2. Diagram of the experimental equipment
图 2. 实验设备图

2.2. 循环老化实验设计

锂离子电池的寿命衰减主要包括日历寿命衰减和循环寿命衰减。由于日历老化试验周期较长，本文

仅关注循环寿命衰减的研究。为了消除日历老化的影响,所有电池需置于同一温控箱内进行试验,并在相同时间结束测试。影响电池寿命的因素较多,包括环境温度、充放电倍率、放电深度(Depth of Discharge, DOD)以及充放电 SOC 范围等。鉴于电池的循环老化过程较为耗时,而研究时间有限,因此需要采取措施加速老化进程。为了加快锂离子电池的寿命衰减实验进程,本文采用高倍率充放电方式进行测试。高倍率充放电能够在较短时间内完成更多的循环次数,从而加速电池内部的物理与化学变化,使衰减趋势更快显现。这不仅能够有效缩短实验周期,提高研究效率,还能在有限的时间内获取更多关于循环老化的关键数据。

实验共选取 156 节 18650 型号的电芯,在严格控制的实验条件下进行充放电循环,以确定每节电池在此模式下能够维持多少个循环。实验采用恒流恒压(CC-CV)模式进行充电,具体流程为:

- 1) 先以 2 C 恒流充电至 4.2 V,再转为恒压充电,直至电流降至 0.05 C 后结束充电。
- 2) 放电过程采用 3 C 恒流模式,放电至 2.75 V 截止。
- 3) 每个充放电循环之间设定 10 分钟间隔,以模拟实际应用环境中的短暂休息阶段,并减少过热影响。
- 4) 每节电芯分别进行 150 至 200 个循环。
- 5) 其中每隔 20 个循环进行一次 RPT 测试。

实验全程记录电池容量变化、内阻演变、电压曲线等关键参数,以便在寿命终止后分析其衰减特性。通过数据分析,我们可以得到不同电芯的寿命差异,从而以便于通过早期分选来判断电池的寿命差异。

2.3. 循环老化结果分析

进行上文的老化实验后得到的结果如图 3 所示,展示了锂离子电池在循环过程中放电容量随循环次数的变化趋势,大多数电池容量随循环次数逐渐衰减,呈现较稳定的下降曲线,而图中用红色虚线圈出的“跳水电池”在短期内容量骤降,表现出早期失效的异常行为。这种现象通常与电池材料缺陷、制造工艺不良或使用条件恶劣有关。该图不仅反映了电池的整体老化规律,也突出了早期识别不良电池的重要性。

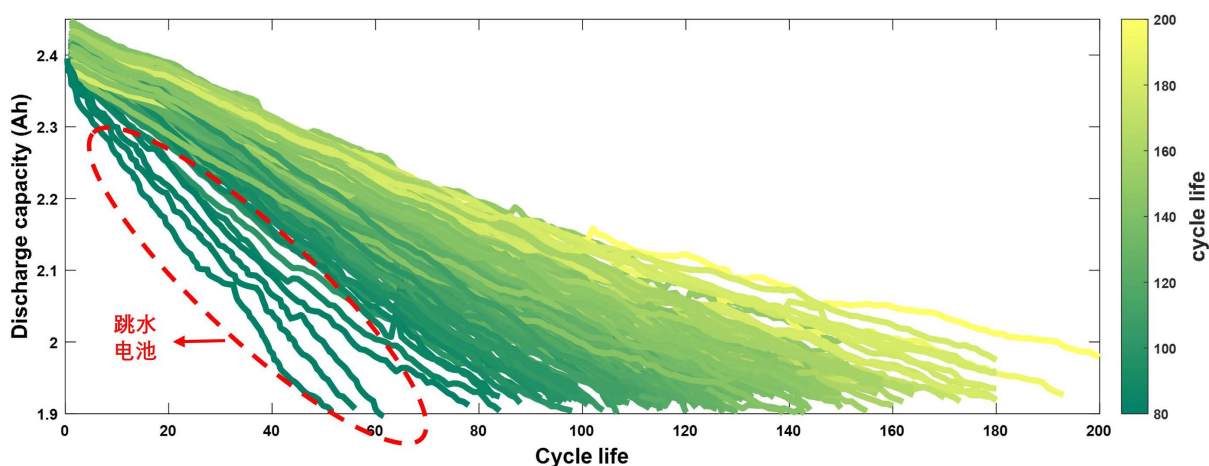


Figure 3. Capacity decay curves of dataset battery samples

图 3. 数据集电池样品的容量衰减曲线

2.4. 数据预处理: 异常数据剔除

实验室电池数据的预处理对于基于神经网络的大模型训练至关重要,直接影响其预测精度和泛化能

力。通过数据清洗、去噪、异常值处理以及归一化等步骤,可提高数据质量并减少测量误差。同时,特征提取与降维优化有助于提高模型的计算效率,而数据增强与不均衡处理能够改善长周期预测的稳定性。此外,不同类型的模型对数据的时序特征和结构化关系有不同的要求,合理的数据预处理可以提升模型对复杂模式的学习能力,从而提高预测的准确性和稳定性。

在电池制造过程中,由于材料均匀性、生产工艺及微观结构等方面的差异,即便是同一批次的电池,其老化过程仍可能表现出不同的退化特性。在相同使用条件下,部分电池可能经历平稳的衰退过程,而另一些电池可能在特定循环阶段出现突发性容量骤降,即“跳水”现象。这种异常衰退模式可能由内阻突增、锂枝晶生长、活性材料脱落或电解液降解等因素导致。若不加以筛选,异常电池的数据可能对模型训练产生干扰,降低预测的准确性和泛化能力。因此,在数据处理阶段,需对实验室采集的电池老化数据进行严格筛选。通过分析容量衰减曲线(图 3),发现有 4 至 5 节电芯在不到 60 个循环时,健康状态就降低到 80%,并且曲线衰减斜率程度远远比正常电芯要高得多。故而剔除异常数据点,以确保用于建模的数据具有代表性和一致性,从而提升后续模型的可靠性和预测性能。剔除完的锂离子电池衰减曲线如图 4 所示。

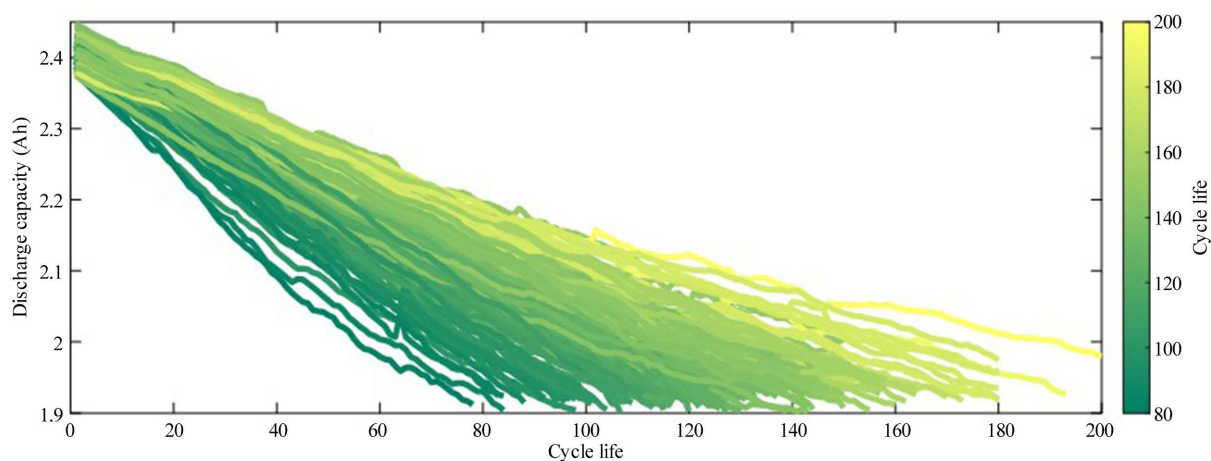


Figure 4. Lithium-ion battery degradation curves after filtering

图 4. 剔除后的锂离子电池衰减曲线

3. 数据增强 - 虚拟电池构建方案

在电池寿命测试和健康状态评估中,实际电池的循环测试数据是非常宝贵的资源。然而,长时间的循环测试往往需要耗费大量的时间和成本。因此,如何高效地利用有限的实验数据以提取有价值的信息,是当前电池研究领域的重要课题。为了解决这一问题,本文提出了一种基于序列采样的虚拟电池构建方案,通过选取实际电池循环数据中的代表性数据段,构建虚拟电池样本,以提升数据利用效率,进而节省实验成本。具体虚拟电池构建方案如图 5 所示。

在电池的实际循环过程中,电池容量会随循环次数的增加而逐渐衰减,假设某节电池的实际循环寿命是在 Cycle 01 至 Cycle End。其中在 n 次循环测试中,电池的容量从初始容量 Q_1 逐渐衰减到 Q_n (通常 n 为 150 次左右),而每个循环测试点的数据能够反映出电池的实际健康状态。为了更有效地利用这些数据,本文引入了“虚拟电池”的概念。具体来说,虚拟电池 VB_i 被定义为一段连续循环数据的集合,其表示为 $VQ_i \sim VQ_i + VQ_n$,其中 i 表示虚拟电池的编号, VQ_i 至 VQ_n 的跨度需要根据电池数据集实际寿命来判断(通常按照整体数据集的评价寿命来定,本章跨度设定为 99)。 n 是随着 i 的变化而变化的。例如,

当 $i = 1$ 、 $n = 100$ 时, VB_1 对应于原始的实际电池数据, 即从初始容量 Q_1 逐渐衰减至 Q_{100} , 代表实际电池的第一个 100 次循环数据段。当 $I = 2$ 、 $n = i + 99 = 101$ 时, 此时 VB_2 对应于原始实际电池数据的第 2 个循环至第 101 个循环, 即从初始容量 Q_2 逐渐衰减至 Q_{101} , 故当 i 等于 1 至某个预定的数值时, VB_i 则代表由原始数据构建出的虚拟电池, 即从第 i 次循环开始的连续 100 次循环数据段, 例如, VB_1 可表示 Q_1 至 Q_{100} , VB_2 则对应 Q_2 至 Q_{101} , 以此类推, 构成了一系列虚拟电池序列 $\{VB_1, VB_2, VB_3, \dots, VB_n\}$ 。

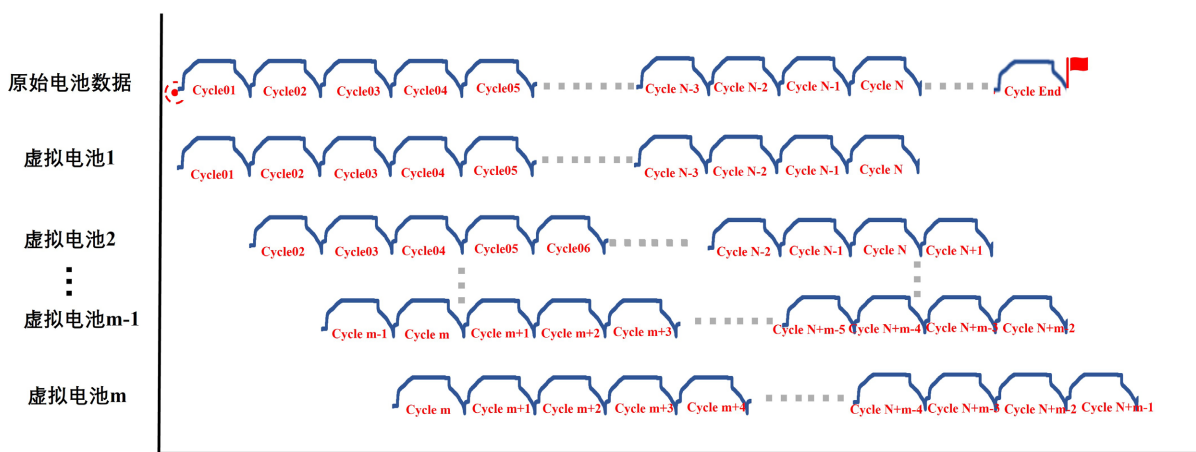


Figure 5. Scheme for constructing virtual batteries using sequence-based sampling

图 5. 基于序列采样的虚拟电池构建方案

通过这种虚拟电池的构建方法, 可以将实际电池的部分循环数据转化为多个虚拟电池样本, 从而在不增加实际实验次数的前提下, 扩展了可用的样本空间。这一方法的核心在于通过滑动窗口的方式从实际电池循环数据中提取具有代表性的连续数据段, 形成多个不同的虚拟电池。每个虚拟电池都保留了实际电池在相应时间段内的性能特征, 能够用于后续的数据分析和模型训练。

虚拟电池序列的构建有助于在电池健康管理和寿命预测中实现数据的高效利用。首先, 这种方法可以显著增加训练样本的数量, 使得基于数据驱动模型能够更好地捕捉电池容量衰减的规律, 提升模型的预测精度。其次, 由于虚拟电池样本是从实际电池数据中直接生成的, 因此能够保留实际电池的物理特性, 避免了单纯数据增强方法可能带来的偏差。此外, 通过选取不同的起始点和窗口大小, 可以灵活调整虚拟电池的数量和覆盖范围, 以适应不同的应用需求和计算资源。故本文提出的基于序列采样的虚拟电池构建方案, 通过构建虚拟电池序列 $\{VB_1, VB_2, VB_3, \dots, VB_n\}$, 实现了对实际电池循环数据的高效利用。这种方法不仅能够增加有效样本数量, 降低实验成本, 还可以为电池健康管理和寿命预测提供更多的参考依据。未来, 进一步的研究可以结合不同类型的电池和应用场景, 对该方案进行优化和验证, 以提升其在实际工程应用中的可行性和适用性。

4. 方法论

本文提出了一种基于 MAE 的全生命周期 EOL (End of Life) 预测方法, 该方法包括四个核心步骤, 如图 6 所示: 数据预处理、数据增强构建虚拟电池样本、MAE 模型预训练和少量带标签数据下的 EOL 预测微调。首先, 在数据预处理阶段, 对数据集进行全面分析, 总结其中的字段特征和统计信息。随后, 进行数据切片和归一化处理, 以确保模型输入的统一性和标准化。接着, 将数据划分为训练集、测试集和验证集, 确保模型的训练与评估具有合理的基础。然后, 通过选取实际电池循环数据中的代表性数据段, 构建虚拟电池样本, 以提升数据利用效率, 进而节省实验成本。在第三步的预训练阶段, 使用无监督学

习的方式对 MAE (Masked Autoencoder)模型进行预训练, 输入数据为未标记的全量数据。这一步的主要目的是通过无监督学习, 使模型初步掌握电池数据中的潜在特征与规律。在此阶段, 模型通过遮蔽部分输入数据并尝试重建来进行训练, 其重建精度被用作评估模型学习效果的主要指标。最后, 利用预训练模型中的编码器部分进行迁移学习, 在微调阶段冻结部分参数, 仅使用少量带标签数据对模型进行微调。通过这种方式, 模型能够更加聚焦于与电池寿命相关的关键特征提取, 最终实现对电池全生命周期的 EOL 精准预测。整个过程既借助了大量无标记数据的潜在信息, 也通过微调增强了模型对特定任务的适应性, 从而实现更高效的 EOL 预测。

通过图 6 所示的流程图说明了基于所提出方法的整体预测过程。该过程包括两个主要部分, 基于电池健康特征提取和基于 MAE 的快速预测全生命周期 EOL。下面将分别阐述每一部分的细节。

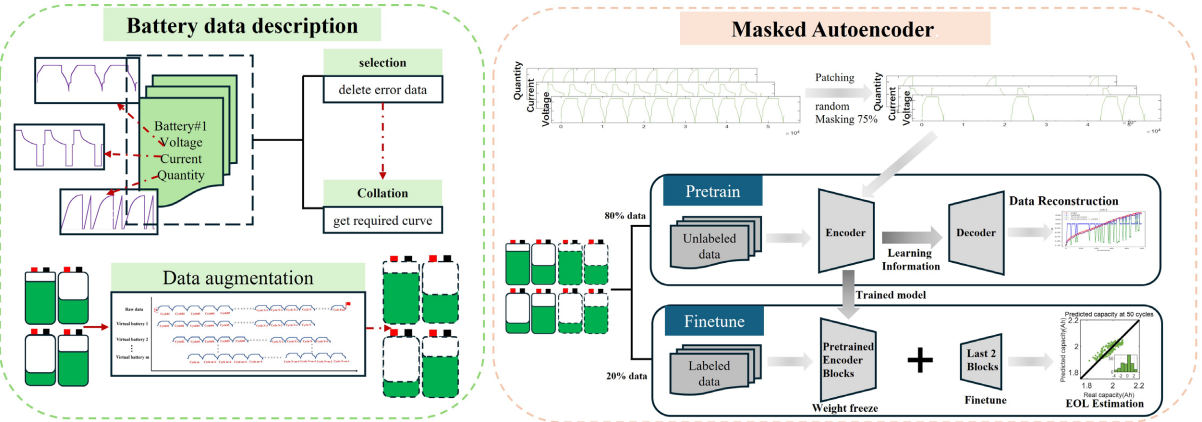


Figure 6. Overall framework of end-of-life (EOL) prediction across the entire battery lifecycle using the MAE approach
图 6. MAE 方法用于电池全生命周期 EOL 预测的总体过程

4.1. 预训练

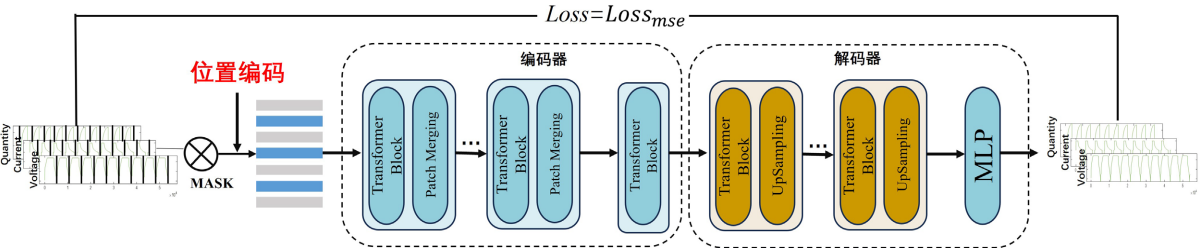


Figure 7. Architecture of the masked pre-training stage
图 7. 掩码预训练阶段框架

在 MAE 的预训练阶段, 模型通过掩码和自监督学习来学习数据曲线的全局表示。具体过程如下: 首先, 输入数据的掩码: 输入的数据首先被分成若干个 patch, 通常大小设置成 50。然后, 从这些块中随机掩盖掉一部分(掩盖 75%的块, 只保留 25%的可见块)。这意味着大多数输入数据曲线信息会被丢失, 模型只能看到少量的数据曲线。编码器(Encoder)处理可见块: 被掩盖的数据曲线不会被直接输入到编码器, 只有保留下来的数据块会被传递给编码器进行处理。编码器采用 ViT (Vision Transformer)结构, 将可见块转换为特征表示。最后通过解码器(Decoder)重建图像: 在编码器生成的特征表示基础上, 解码器的任务是重建原始的输入数据, 包括那些被掩盖掉的数据块。这部分需要解码器根据少量的可见信息来预测出丢失的数据块, 模型的学习目标是 최소화重建数据和原始数据之间的差异。

在这个阶段，模型没有使用任何标签信息，完全通过图像数据的自监督进行训练。这种方法可以帮助模型学习到丰富的、与上下文相关的特征表示。MAE 掩码预训练框架如图 7 所示，重构结果如图 8 所示。

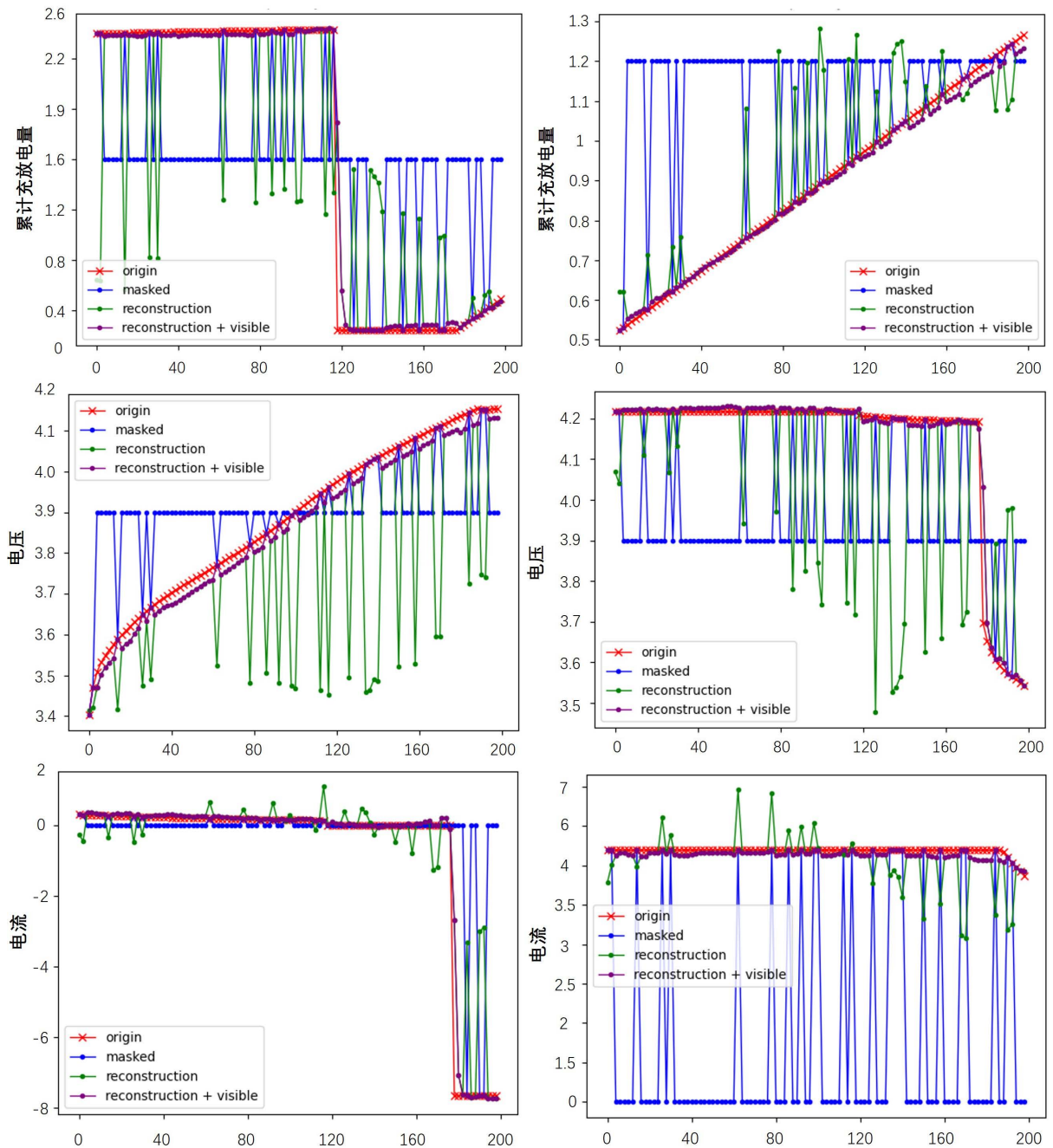


Figure 8. Reconstruction results

图 8. 重构结果图

4.2. 微调阶段

微调阶段是为了让模型在特定的下游任务中表现更好。这一阶段与传统的有监督学习类似，步骤如下：输入完整图像：微调时，输入的数据不再掩盖块，而是完整地输入到模型中。这是因为微调阶段的

目标是解决具体的任务，不再需要掩码重建。使用预训练的编码器：微调阶段使用在预训练阶段得到的编码器，并且保留了其学到的权重。这些预训练的权重有助于模型快速适应下游任务。由于在预训练阶段模型已经学会了如何有效地提取数据的全局特征，微调时模型能够更加快速、精准地针对任务进行学习。最后进行有监督训练：在微调阶段，模型会使用标注数据进行有监督的训练。例如，在图像分类任务中，模型通过分类头(Classifier Head)输出类别标签，并根据预测结果与真实标签之间的误差来更新模型的权重。这种微调方式可以有效利用预训练时学到的知识，加速训练过程并提高性能。

5. 结果与讨论

在本节中，设计并进行了一系列计算实验，以证明所提出方法在快速电池分类方面的优势。计算研究中使用了来自自测数据集的总计 154 个原始电池样本。

这些电池样本具有不同的循环寿命、不同的初始容量、不同的 EOL。预训练集和微调集的划分采用分层随机抽样方法。电池样本按 70%预训练集和 30%微调集的比例进行划分，微调集中又设置随机种子，其中 70%用于训练，15%用于测试，15%用于验证。

在电池质量分类中，由于考虑到电池单体组成电池包后，电池包在相同使用条件和相同循环圈数时内部锂离子电池单体的老化速率不同，会出现锂离子电池老化快慢不均的情况，这也是导致电池包衰老加快的重要因素。故我们从全生命周期的 EOL 出发去进行电池早期分选，对传统只预测最终 EOL 来进行早期分选的方法进行优化补充。由于如图 3 所示，前 50 个循环锂离子电池并未出现明显衰老异常的情况。故考虑了从 50 次循环后进行全生命周期 EOL 预测。所以电池样本都在 50 个循环之后(50—100，每 10 个循环为间隔)进行 6 次不同循环时的容量预测。在此设置中，使用四个度量来评估所提出的方法的性能，百分比误差(PE)、均方误差(MSE)、均方根误差(RMSE)、平均绝对百分比误差(MAPE)，如以下公式所示：

$$PE = \frac{\hat{y} - y}{y} \times 100\%$$

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2}$$

$$MSE = \frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2$$

$$MAPE = \frac{100\%}{m} \sum_{i=1}^m \left| \frac{\hat{y}_i - y_i}{y_i} \right|$$

其中 m 、 y_i 、 $\hat{y}_i$ 分别表示样本数、真实值、预测值。

5.1. 参数量选择的影响

不同数量参数早期循环电池数据包含不同数量的电池退化信息，这可能影响方法的性能。为了探索不同参数量对电池全生命周期 EOL 预测的影响，使用来自 1 个参数(电压)、3 个参数(电压、电流、SOC)、5 个参数(电压、电流、SOC、能量、dQ/dV)的原始电池数据作为方法输入进行比较实验。首先研究了直接预测所有样本电池的第 100 个循环容量。通过预测第 100 个循环时的容量来确定最终参数量的选择。

图 9 总结了此次对比实验的结果。

如图 9 所示，当考虑来自前 5 个参数(电压、电流、SOC、能量、dQ/dV)的电池数据时，获得了最

佳预测结果。三项指标 MSE、RMSE、MAPE 的结果最好, 分别为 0.000 84%、0.028 987 1%和 1.212 041%。在使用 3 个参数(电压、电流、SOC)的时候, 三项指标的结果分别为 0.000 91%、0.0298 274%和 1.220 83%。在仅仅使用一个参数(电压)的情况下, 预测性能也是有希望的。三项指标的结果分别为 0.001 24%、0.038 987%、1.912 041%。可以观察到模型性能随着输入参数数量的增加而逐渐改善。例如, RMSE 从 0.038 987% (1 个参数)降低到 0.029 827 4% (3 个参数)和 0.028 987 1% (5 个参数)。其他两个指标也呈现出类似的降低模式。这种性能改进可以通过在考虑更多参数时利用来自原始电池数据的更多空间和时间信息来解释。另一方面, 在使用 3 个参数和 5 个参数时, 预测性能没有太大的提高, 甚至随着参数数量的增减而变得有点接近。例如, 通过 MAPE 观察到 1.212 041% (5 个参数)和 1.220 83% (3 个参数)几乎没有差别。所以从而猜想, 扩展参数数量不会带来太多性能增益, 相反会增加测试时间和计算成本, 这是不必要的。

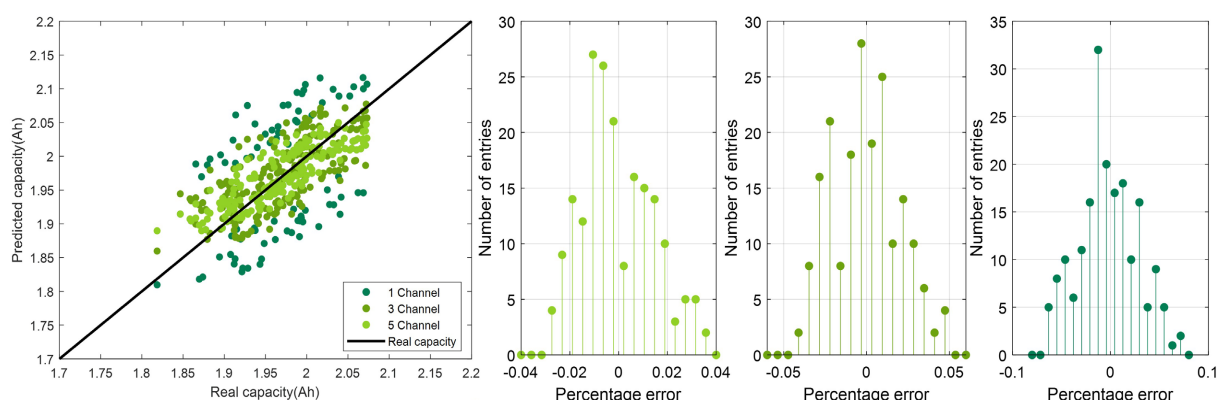


Figure 9. Predicted capacity for 100 cycles and prediction percentage errors with different parameter sets

图 9. 同参数数量时预测 100 个循环容量的预测结果及预测百分比误差

通过对图 9 的分析可以看出, 从横坐标的 PE 指标来看, 使用不同数量的参数对预测结果产生了显著影响。当使用 1 个参数(电压)时, PE 值的范围在-0.1 到 0.1 之间; 使用 3 个参数(电压、电流、SOC)时, PE 值的范围缩小至-0.04 到 0.04; 而使用 5 个参数(电压、电流、SOC、能量、dQ/dV)时, PE 值的范围为-0.05 到 0.05。进一步从纵坐标的占比数量来看, 使用 1 个参数(电压)时, PE 值在-0.1 和 0.1 附近的电池数量占比较大。而使用 3 个参数(电压、电流、SOC)和 5 个参数(电压、电流、SOC、能量、dQ/dV)时, PE 值分别集中在-0.04 和 0.04、-0.05 和 0.05 附近, 且电池数量占比较少。

综合分析可知, 尽管使用 1 个参数(电压)也能得到相对良好的预测效果, 但相比之下, 使用 3 个参数(电压、电流、SOC)能够在保证精度的同时, 更有效地进行电池分类, 且预测速度较快。因此, 在后续分析中, 除非另有特别说明, 默认使用包含 3 个参数(电压、电流、SOC)的电池数据进行研究和验证。

5.2. 预测不同循环周期容量结果及验证

通过以上对于超参数规模、循环周期数、输入参数数量的选择, 最终确定使用 tiny 的超参数规模、前 3 个循环周期以及参数数量(电压、电流、SOC)来预测全生命周期 EOL。分别预测了 50、60、70、80、90、100 个循环周期的容量。结果如图 10 所示。

从图 10 可以观察到, 在预测全生命周期 EOL 的时候, 三个判定指标上都呈现出较好的精度结果。同时为了验证我们模型的准确性, 选择了额外的 3 节电池作为验证, 验证结果如图 10 所示。结果显示在预测不同循环周期时最大 RMSE 不超过 0.04。

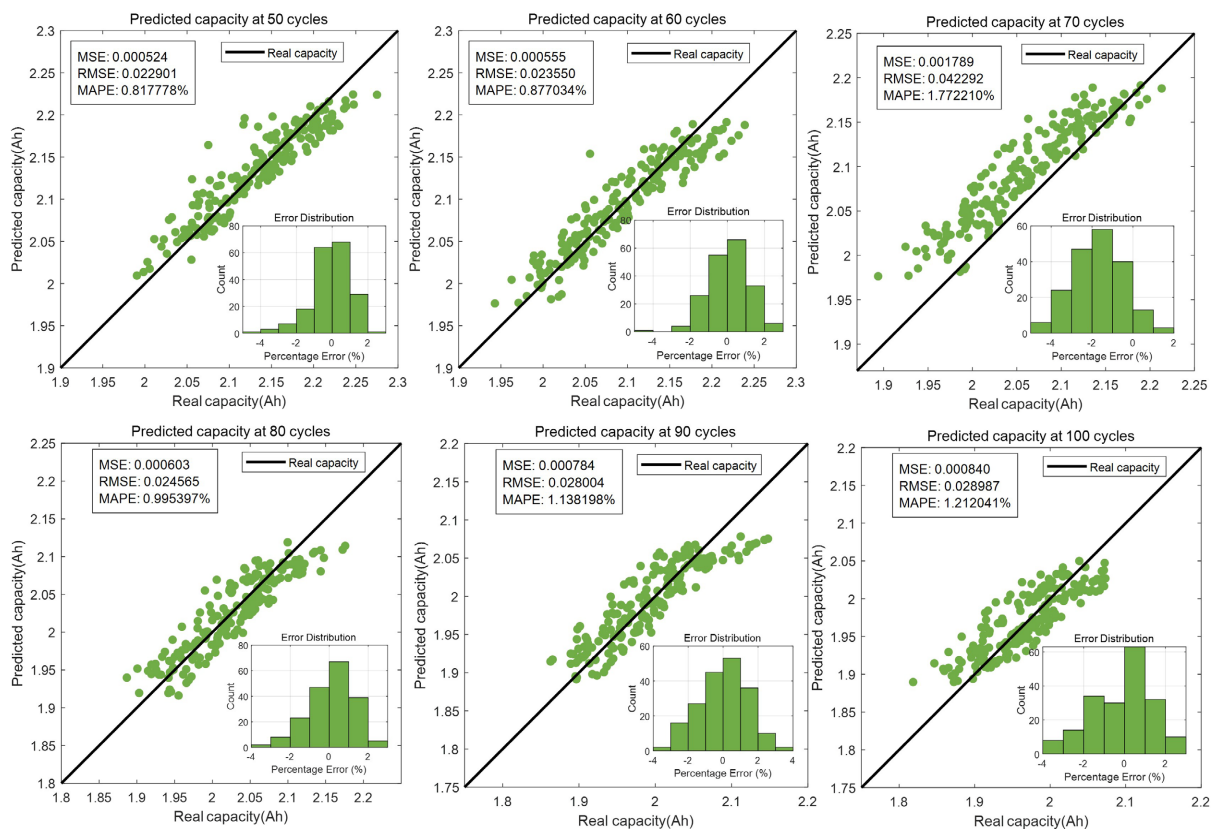


Figure 10. Performance evaluation of predictions across different cycling periods
图 10. 预测不同循环周期的性能分析

5.3. MIT 数据集验证

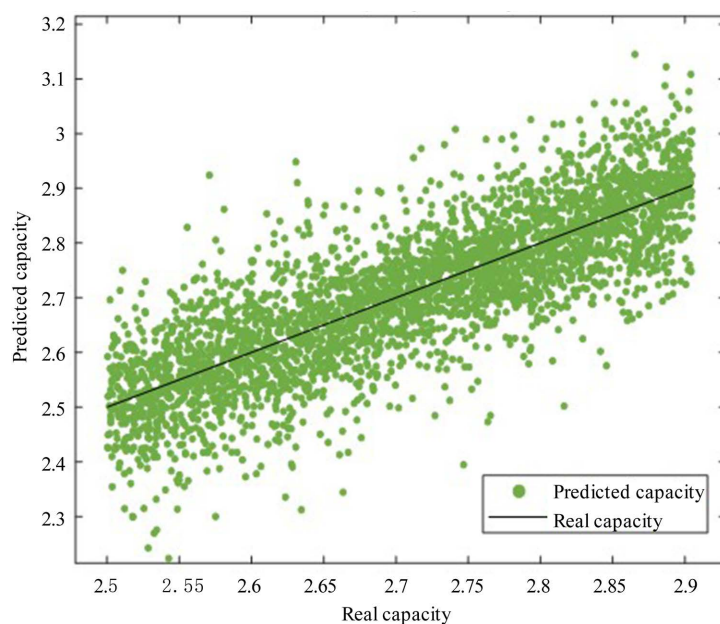


Figure 11. Prediction results on the MIT dataset
图 11. MIT 数据集的预测结果

为了更加全面地评估所提出模型的性能与泛化能力,本文在完成基础训练及初步验证之后,进一步引入了一个独立于训练过程的 MIT 数据集进行验证分析。该数据集与原始训练集在来源、分布特征或采集条件上存在一定差异,能够更真实地反映模型在不同数据环境下的适应性和稳定性。通过在该数据集上进行验证,我们不仅考察了模型在非同源数据上的预测能力,也从侧面检验了模型是否存在过拟合现象。结果如图 11 所示。

6. 结论

本研究的主要贡献总结如下:

- 1) 提出了一种基于序列采样的虚拟电池构建方案及基于半监督学习的全生命周期 EOL 预测方法。该方法利用无标签数据进行预训练,并结合少量有标签数据进行微调,使得模型能够有效地学习 EOL 预测的关键特征,最终达到较高的预测精度。
- 2) 利用前三个循环数据进行 EOL 预测。相较于传统方法依赖全生命周期数据进行 EOL 预测,本研究的方法基于早期循环数据即可做出可靠的预测,从而实现电池的早期分选,优化电池筛选流程。
- 3) 优化电池分选流程,减少不良品率。由于本方法能够在电池早期阶段进行 EOL 预测,因此能够在成组之前筛选出潜在的不良电池,从而减少后续电池包的不良率,提高电池系统的整体性能。
- 4) 提高模型的训练效率和计算资源利用率。通过实验对比分析,本研究优化了输入数据参数量和超参数配置,使得模型在保证预测精度的同时,大幅降低了训练时间和计算资源需求。

参考文献

- [1] Lai, X., Chen, Q., Tang, X., Zhou, Y., Gao, F., Guo, Y., *et al.* (2022) Critical Review of Life Cycle Assessment of Lithium-Ion Batteries for Electric Vehicles: A Lifespan Perspective. *E Transportation*, **12**, Article 100169. <https://doi.org/10.1016/j.etrans.2022.100169>
- [2] Su, L., Wu, M., Li, Z. and Zhang, J. (2021) Cycle Life Prediction of Lithium-Ion Batteries Based on Data-Driven Methods. *E Transportation*, **10**, Article 100137. <https://doi.org/10.1016/j.etrans.2021.100137>
- [3] Tomaszewska, A., Chu, Z., Feng, X., O’Kane, S., Liu, X., Chen, J., *et al.* (2019) Lithium-Ion Battery Fast Charging: A Review. *E Transportation*, **1**, Article 100011. <https://doi.org/10.1016/j.etrans.2019.100011>
- [4] Attia, P.M., Grover, A., Jin, N., Severson, K.A., Markov, T.M., Liao, Y., *et al.* (2020) Closed-Loop Optimization of Fast-Charging Protocols for Batteries with Machine Learning. *Nature*, **578**, 397-402. <https://doi.org/10.1038/s41586-020-1994-5>
- [5] Severson, K.A., Attia, P.M., Jin, N., Perkins, N., Jiang, B., Yang, Z., *et al.* (2019) Data-Driven Prediction of Battery Cycle Life before Capacity Degradation. *Nature Energy*, **4**, 383-391. <https://doi.org/10.1038/s41560-019-0356-8>
- [6] Fei, Z., Yang, F., Tsui, K., Li, L. and Zhang, Z. (2021) Early Prediction of Battery Lifetime via a Machine Learning Based Framework. *Energy*, **225**, Article 120205. <https://doi.org/10.1016/j.energy.2021.120205>
- [7] Tian, J., Wang, Y., Liu, C. and Chen, Z. (2020) Consistency Evaluation and Cluster Analysis for Lithium-Ion Battery Pack in Electric Vehicles. *Energy*, **194**, Article 116944. <https://doi.org/10.1016/j.energy.2020.116944>
- [8] Lu, Y., Li, K., Han, X., Feng, X., Chu, Z., Lu, L., *et al.* (2020) A Method of Cell-to-Cell Variation Evaluation for Battery Packs in Electric Vehicles with Charging Cloud Data. *E Transportation*, **6**, Article 100077. <https://doi.org/10.1016/j.etrans.2020.100077>
- [9] Cheng, M., Sun, H., Wei, G., Zhou, G. and Zhang, X. (2022) A Sustainable Framework for the Second-Life Battery Ecosystem Based on Blockchain. *E Transportation*, **14**, Article 100206. <https://doi.org/10.1016/j.etrans.2022.100206>
- [10] Huang, Z., Xie, Z., Zhang, C., Chan, S.H., Milewski, J., Xie, Y., *et al.* (2019) Modeling and Multi-Objective Optimization of a Stand-Alone PV-Hydrogen-Retired EV Battery Hybrid Energy System. *Energy Conversion and Management*, **181**, 80-92. <https://doi.org/10.1016/j.enconman.2018.11.079>