

부분 방전 구간에서의 배터리 열화 합성데이터 생성을 위한 물리-데이터 기반 하이브리드 프레임워크

최승민*, 김광섭**, 최현준***¹, 박범용***², 정유철****

A Physics-Guided Hybrid Framework for Synthetic Battery Degradation Data Generation under Partial Depth-of-Discharge Conditions

Seungmin Choi*, KwangSeop Kim**, HyunJun Choi***¹, BumYong Park***², and Yuchul Jung****

본 연구는 2026년 정부 재원으로 국방과학연구소의 지원(UE251068HD)과 2026년도 교육부 및 경상북도의 재원으로 경북RISE센터의 지원을 받아 수행된 지역혁신중심 대학지원체계(RISE)-(특화산업 Scale-up)의 결과입니다(2026-rise-15-105)

요 약

제한된 DoD(Depth of Discharge) 조건에서 획득한 배터리 열화 데이터는 실제 운용 환경을 반영하지만, 획득 비용이 크고 분포 범위가 제한되어 학습용 데이터셋 구축이 어렵다. 본 연구는 이를 보완하기 위해 정규화, 물리 파라미터 식별, 물리 기반 시드 생성, 대리모델 기반 대량 생성, 요약 공간 잔차 보정, 품질 관리 게이트로 구성된 물리-데이터 하이브리드 합성 데이터 생성 프레임워크를 제안한다. 제안 프레임워크는 SoC(State of Charge) 40%~85%의 제한 DoD 데이터를 대상으로 그룹별 분포 정합성과 물리 일관성을 검증하고, 품질 기준을 만족한 합성 데이터만 최종 학습 데이터셋에 포함한다. 내부 실험에서는 LG M50L 셀 6개와 4,027개 실측 cycle 요약을 기반으로 375,207개의 제한 DoD 합성 데이터를 구축했으며, 외부 공개 데이터셋에 단계적으로 적용해 외부 적용 가능성도 확인하였다.

Abstract

Battery degradation data acquired under restricted Depth-of-Discharge (DoD) conditions reflect real-world operating environments; however, constructing training datasets remains difficult because of high acquisition costs and limited distributional coverage. To address these limitations, this study proposes a physics - data hybrid synthetic data generation framework comprising normalization, physical parameter identification, physics-based seed generation, surrogate-model-based large-scale generation, residual correction in a summary space, and quality-control gates. The proposed framework evaluates both group-level distributional alignment and physical consistency using data acquired within a restricted State-of-Charge (SoC) range of 40% - 85%, and includes only synthetic data satisfying the quality criteria in the final training dataset. In the internal experiments, 375,207 restricted-DoD synthetic samples were generated from six LG M50L cells and 4,027 measured cycle summaries. The framework was also applied progressively to external public datasets to assess its broader applicability.

Keywords

battery degradation, DoD window constraints, physics-based parameter identification, synthetic data generation

* 국립금오공과대학교 컴퓨터·AI융합공학과
 - ORCID: <https://orcid.org/0009-0007-8243-9648>
 ** 국방과학연구소 선임연구원
 - ORCID: <https://orcid.org/0000-0002-2130-5925>
 *** 국립금오공과대학교 전자공학부(***² 공동교신저자)
 - ORCID¹: <http://orcid.org/0000-0002-8479-9256>
 - ORCID²: <https://orcid.org/0000-0001-9490-0365>
 **** 국립금오공과대학교 인공지능공학과 부교수(공동교신저자)
 - ORCID: <http://orcid.org/0000-0002-8871-1979>

• Received: Mar. 20, 2026 Revised: Apr. 16, 2026, Accepted: Apr. 19, 2026
 • Co-corresponding Author1: Bumyong Park
 Dept. of Electronic Engineering, Kumoh National Institute of Technology,
 61 Daehak-ro (yangho-dong), Gumi, Gyeongbuk, [39177] Korea
 Tel.: +82-478-7438, Email: bumyong.park@kumoh.ac.kr
 • Co-corresponding Author2: Yuchul Jung
 Dept. of AI Engineering, Kumoh National Institute of Technology, 61
 Daehak-ro (yangho-dong), Gumi, Gyeongbuk, [39177] Korea
 Tel.: +82-478-7436, Email: jyc@kumoh.ac.kr

1. 서 론

리튬이온 배터리의 열화는 충·방전 범위, 전압 상한, 온도, 휴지 시간 등 운전 조건과 내부 반응의 상호 작용에 의해 결정된다[1][2]. 특히 부분 충·방전으로 운전되는 특정 DoD(Depth-of-Discharge) window는 전주기 운전과 다른 열화 양상을 보일 수 있으며, 부분 DoD cycling은 수명에 실질적인 영향을 줄 수 있다[3]. 또한 제한된 window에서 관측된 방전용량은 절대 SoH(State-of-Health)를 직접 반영하지 않을 수 있으므로, 해당 조건의 열화 해석에는 별도의 기준 축이 필요하다[4].

이러한 데이터 부족 문제를 보완하기 위한 접근으로는 합성 데이터 생성과 데이터 증강 기반 접근이 제안되어왔다[5]-[8]. 해당 과정에서 계산 비용을 낮추기 위한 대리모델 기반 시뮬레이션 연구도 같이 보고되는데[9][10], 기존의 순수 데이터 기반 생성은 표본 수가 적을 때 조건별 분포와 열화 추세를 안정적으로 재현하기 어렵고, 순수 물리 기반의 시뮬레이션은 실제 계측 노이즈, 운전 변동성, 조건 전이 구조를 충분히 반영하지 못하는 경우가 많다[11]. 따라서 실측 데이터의 관측 구조를 유지하면서 물리적 일관성까지 확보할 수 있는 물리-데이터 하이브리드 접근이 필요하다[12].

본 논문은 SoC(State-of-Charge) 40%~85%의 부분 충·방전 조건에서 제한적으로 획득되는 배터리 열화 데이터를 보완하기 위해, 학습용 데이터셋 구축을 위한 합성 데이터 생성 프레임워크를 제안한다.

제안 프레임워크는 그림 1과 같이 다섯 단계로 구성된다. Stage A에서는 실측 데이터 정규화와 참조 SoH 축 구성을 수행하고, Stage B에서는 PyBaMM[13]과 PyBOP[14]을 이용하여 물리 파라미터를 식별한다. Stage C에서는 물리 모델 기반 시드 생성과 대리모델 기반 대량 생성을 수행하며[9][10][15][16], Stage D에서는 요약 공간 잔차 보정을 적용한다. 마지막으로 Stage E에서는 품질 관리 게이트를 통해 최종 합성 데이터를 선별한다.

본 프레임워크의 핵심은 제한된 실측 데이터를 복제하는 것이 아니라, 물리적으로 해석이 가능한 파라미터의 공간과 열화 축을 복원 및 추정된 뒤, 이를 기반으로 목표 조건에 특화된 합성 데이터를 생성하는 것이다. LG사 M50L 배터리기반의 실측 데이터 6개 셀과 4,027 cycle 요약 데이터를 입력으로 하여, 최종적으로 375,207개의 합성 데이터를 생성한다.

본 논문의 구성은 다음과 같다. 2장에서는 제한 DoD 데이터셋의 구성과 Stage A부터 Stage E까지의 세부 설계를 설명한다. 3장에서는 그룹별 실측 데이터 특성, 품질 관리 게이트 결과, Stage D의 잔차 보정 효과, 그리고 외부 데이터셋 적용 결과 및 비교 실험 결과를 제시한다. 마지막으로 4장에서는 본 연구의 의의와 한계, 향후 확장 방향을 논의한다.

본 연구의 기여는 다음과 같다. 첫째, 제한 DoD 조건에서 관측되는 부분 방전용량을 절대 SoH와 직접 등치하지 않고, RPT anchor와 cycle-level 정렬을 결합한 참조 SoH 축 구성 절차를 제시하였다.

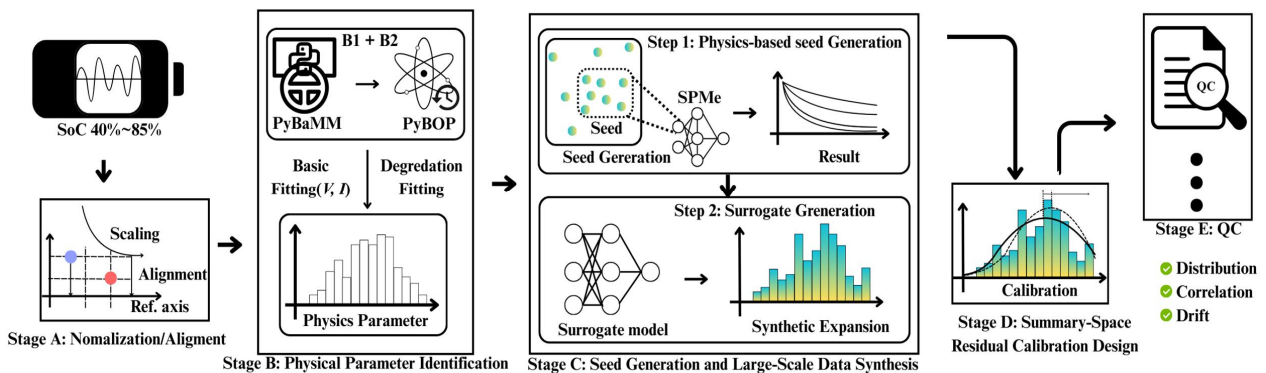


그림 1. 제안 프레임워크 개요도

Fig. 1. Overview of the proposed framework

이를 통해 제한 DoD 환경에서도 후속 파라미터 식별과 합성 데이터 생성에 사용할 수 있는 일관된 기준 좌표계를 마련하였다. 둘째, 전기·열 파라미터와 열화 파라미터를 분리하여 단계적으로 식별하고, 이를 SPM_e 기반 시드 생성, 대리모델 기반 확장, 요약 공간 잔차 보정과 연결한 물리-데이터 하이브리드 합성 파이프라인을 제안하였다.

마지막으로 그룹별 분포를 유지한 상태에서 ROC-AUC, KS 통계량, 상관 구조 차이, SoH 단조성 위반률, 재시물레이션 drift를 함께 사용하는 품질 관리 게이트를 도입함으로써, 분포 정합성과 물리 일관성을 동시에 점검하는 검증 절차를 제시하였다. 본 연구는 제한 DoD 환경에서 확보가 어려운 열화 데이터를 학습 가능한 규모로 확장하면서도 물리 해석 가능성과 데이터 활용성을 함께 유지하려는 실용적 프레임워크를 제안한다.

II. 데이터셋 및 실험 설계

2.1 제한 DoD 데이터셋 구성

본 연구에서는 LG M50L 셀을 대상으로 SoC 40%~85% 범위의 부분 충·방전 조건에서 획득한 제한 DoD 열화 데이터를 사용하였다. 해당 데이터는 다채널 배터리 시험기에서 수행한 반복 충·방전 실험으로부터 획득하였으며, 상한 전압 4.2V, 하한 전압 2.5V, 주 반복 구간 전류 크기 약 2.4A, CV 컷오프 0.24A, nominal rest 1,800s의 공통된 설정을 가진다. 그룹 A는 4개 셀로 구성되며 각 셀의 관측 길이는 431, 822, 859, 813 cycle이었다. 그룹 B는 2개 셀로 구성되며 각 셀의 관측 길이는 598, 589 cycle이었다. 두 그룹 모두 진단 anchor는 50-cycle 간격으로 설정하였다.

표 1. 제한 DoD 데이터셋 개수 정리

Table 1. Number of constrained DoD datasets

Category	Group		Final synthetic sata
	A	B	
Cell	4	2	6
Cycle summary	2,864	1,163	4,027
Accepted synthetic rows	187,929	187,278	375,207

그룹 B는 그룹 A와 동일 계열의 실험에서 획득되었으나, 연속적인 제한 DoD 세그먼트가 충분하지 않았다. 이에 따라 전체 실험 데이터 중 연구 목적에 부합하는 cycle만을 선별하고, 분포 확보에 필요한 최소 길이를 만족하는 실험만 포함하였다. 그 결과 본 연구에서 사용한 제한 DoD 데이터셋은 표 1과 같이 총 6개 셀, 4,027개의 cycle 요약, 148개의 진단 기준점으로 구성되었다.

메인 데이터셋에 포함된 두 실험군은 운전 특성에서도 차이를 보였는데, 한 집단의 경우, 방전용량과 운전시간의 중앙값이 각각 약 1.02Ah, 1.14h이고, 다른 집단은 약 1.49Ah, 2.32h다. 따라서 동일한 제한 DoD 조건이라도 실제 window의 크기와 지속 시간은 단일 평균 분포로 설명되기 어렵다. 본 논문은 이러한 구조를 반영하기 위해, 모든 데이터를 단순 평균화하지 않고 실측 조건의 분포 구조를 유지하여, 이후의 절차를 진행한다.

2.2 Stage A: 데이터 정렬 및 참조 SoH 축

Stage A의 목적은 실측 실험 로그와 기록 정보를 일관된 기준 좌표로 정렬하는 것에 있다. 먼저, 실측 충·방전 데이터를 cycle 단위의 요약으로 변환하고 전압 V , 전류 I , 시간 t , 방전용량 Q_{dis} , 온도 T 를 기반으로 운전 조건을 분류한다. 이후 진단 cycle을 기준으로 참조 SoH 축을 구성하고, 일반 cycle 축으로 보간하여 후속 단계의 기준 축으로 사용한다. 이러한 절차는 제한 DoD 조건에서 관측되는 방전량이 절대 SoH를 직접 나타내지 않기에 필요하다. 부분 충·방전 데이터의 방전용량은 실제 열화와 더불어 window의 크기 영향을 함께 반영하므로, 그대로 SoH를 해석할 경우, 열화 축의 왜곡이 발생할 수 있다[3][4].

이를 해결하기 위해, 주기적으로 삽입되는 RPT (Reference Performance Test) 식별 cycle에서 얻은 기준점을 바탕으로 참조 SoH를 구성하고, non-RPT의 경우, RPT anchor에서 계산한 SoH를 보간 및 전파하여 참조 SoH 축을 구성한다. 이후 단계의 파라미터 식별과 합성 데이터 보정의 기준으로 사용한다. 또한, 방전량이 비정상적으로 작거나 직전 cycle 대

비 급격한 변화가 이뤄지는 경우, 이상치 후보로 기록하여 이후에 미치는 영향을 최소화한다.

제한 DoD 여부의 경우 파일 단위에서 RPT cycle과 non-RPT cycle의 방전용량 중앙값 비율로 판단하여 식 (1)과 같이 정의한다.

$$r_{\text{window}} = \frac{\text{median}(Q_{\text{dis}}^{\text{non-RPT}})}{\text{median}(Q_{\text{dis}}^{\text{RPT}})} \quad (1)$$

식 (1)에서 $Q_{\text{dis}}^{\text{non-RPT}}$ 는 일반적인 부분 충·방전 cycle의 용량이고, $Q_{\text{dis}}^{\text{RPT}}$ 는 같은 파일 내, 진단 cycle 기준으로 정의된 참조 용량이다. 마지막으로 $r_{\text{window}} < 0.60$ 을 제한 DoD 조건 판정의 기준으로 사용한다.

2.3 Stage B: 물리 파라미터 추정

Stage B에서는 제한 DoD 조건에서 관측된 전압, 온도, 방전용량 데이터를 바탕으로, 배터리 운전을 설명할 수 있는 전기·열 및 열화 관련 물리 파라미터를 단계적으로 식별한다. 파라미터 식별은 B1과 B2의 두 단계로 진행한다.

먼저 초기 동특성과 장기 열화 거동을 한 번에 추정할 경우, 최적화가 불안정해질 가능성이 존재하여, B1 단계에서는 접촉저항, 열전달계수, 공칭 용량, 음극 초기 농도와 같이 초기 응답을 좌우하는 전기·열 파라미터를 추정한다. 이후 B2 단계에서는 열화 메커니즘이 포함된 전기화학 모델을 사용해, 선택된 진단 cycle의 전압, 온도, 방전용량 응답을 잘 재현할 수 있도록 SEI(Solid Electrolyte Interphase) 성장, 리튬도금, 활성물질 손실 및 음극 확산에 대응하는 유효 파라미터를 간접 추정하였다.

이를 위해, 파라미터 식별에는 PyBaMM(Python Battery Mathematical Modeling)[13]과 PyBOP(Python Battery Optimization and Parameterization)[14]을 사용하였다. PyBaMM은 배터리의 전기화학 및 열 응답을 계산하는 전방 모델을 위해 사용하였고, PyBOP은 실측 데이터와 모델 출력 사이의 오차를 최소화하도록 파라미터를 추정 및 조정하기 위한 최적화 도구로 활용되었는데, 전기화학 모델의 응답이 실제

측정값을 재현하도록 물리 파라미터를 역으로 추정하게 된다.

각 단계의 목적 함수는 실측값과 시뮬레이션 결과 사이의 오차를 바탕으로 정의하였다. B1 단계에서는 전압과 평균 셀 온도에 대한 정규화된 RMSE의 가중합을 사용하였고, B2 단계에서는 여기에 방전용량 항을 추가하여 정의하였다. 즉, B1은 초기 전기·열 응답의 정합에 초점을 두고, B2는 열화에는 따른 용량 변화까지 함께 반영하도록 확장된 목적 함수를 사용한다. 각 단계의 목적 함수는 식 (2)와 식 (3)과 같다.

$$J_{B1} = \frac{\text{RMSE}_V}{0.05} + \frac{\text{RMSE}_T}{2.0} \quad (2)$$

$$J_{B2} = \frac{\text{RMSE}_V}{0.05} + \frac{\text{RMSE}_T}{2.0} + \frac{\text{RMSE}_Q}{1.0} \quad (3)$$

최적화는 모두 다중 시작점 방식으로 수행하며, B1과 B2의 시작점은 각각 6개, 4개로 진행한다. 해당 파라미터 표본은 B1에서 24건, 12건의 적합을, B2에서는 16건, 12건을 성공하였으나, B1과 B2의 적합 오차 합을 기준으로 상위 표본 선발건은 각각 10개, 5개가 최종 파라미터 표본 집합이 구축됐다. 최종적으로 15개의 통합 물리 파라미터 표본 집합을 구축하여, 해당 분포를 Stage C의 합성 생성에 대한 입력으로 활용한다.

2.4 Stage C: 시드 생성 및 데이터 대량 생성

Stage C에서는 Stage A에서 정의한 운전 조건 구간과 Stage B에서 추정한 물리 파라미터 분포를 바탕으로, 제한 DoD 조건에서의 합성 데이터를 생성한다. 이 단계는 실측 데이터가 가지는 분포적 특성과 물리적 타당성을 최대한 유지하면서 부족한 데이터를 충분한 규모로 확장하는 데 초점을 둔다. 이를 위해 생성 과정은 물리 기반 시드 생성과 대리 모델 기반 확장의 두 단계로 나누어 구성하였다.

먼저 첫 번째 단계에서는 SPM(Single Particle Model with electrolyte)을 이용해 물리 기반 시드 데이터를 생성한다. 이때 SPM 시뮬레이션에는 Stage

B에서 식별한 물리 파라미터와 Stage A에서 정리한 운전 스케줄을 함께 반영하였으며, 전기화학적 제약을 만족하는 cycle 요약량 수준의 시드를 얻을 수 있다. 이러한 시드는 물리 모델에 기반한 기준 데이터를 제공한다는 점에서 의미가 크지만, SPMe 자체의 계산 비용이 높아 대규모 생성에 활용하기에는 한계가 있다.

이 한계를 보완하기 위해 두 번째 단계에서는 대리모델(Surrogate model)로 SPMe로부터 얻은 시드를 학습한다[9][10]. 입력 조건과 cycle 요약 사이의 관계를 근사하도록 설계하기 위해, SPMe와 동일한 입력 변수 집합을 사용하여 운전 조건과 물리 파라미터를 입력으로 받고, 이에 대응하는 cycle 요약을 예측하도록 MLP(Multilayer Perceptron) 기반 모델 학습을 진행한다. 즉, SPMe가 물리적으로 신뢰할 수 있는 소규모 시드를 만드는 역할을 한다면, 대리모델은 그 관계를 학습해, 보다 효율적으로 대규모 합성 데이터를 생성하는 역할을 한다.

대리모델의 경우, 조건 식별자는 8차원 임베딩으로 256-256-128의 은닉층을 가졌으며, 각 은닉층은 GELU 활성화 함수를 적용했다. 은닉층 뒤에 과적합 완화를 위한 dropout 0.1로 구성하였으며, 학습은 AdamW 옵티마이저로 learning rate = 3×10^{-4} , weight decay = 1×10^{-4} , batchsize = 4096, maximum epoch = 80, validation ratio = 0.1로 설정하였으며, 손실 함수는 가중 Smooth L1(Huber) loss로 설정했다.

합성 데이터의 생성 규모는 임의의 고정값으로 정하지 않고, 실측 데이터에서 관찰된 유효 셀 수와 수명 길이를 반영해 자동으로 조정하였다. 또한 제한 DoD 조건에서 물리적 타당성을 유지하기 위해 몇 가지 제약 조건도 함께 적용하였다. 최소 방전 용량은 3.0 Ah, 용량 비율은 0.55 이상, 최대 접촉 저항은 0.015, 최대 초기 전압 강하는 1.20 V로 설정하였으며, 이러한 임계값은 실측 데이터에서 관찰된 경험적 범위를 바탕으로 정하였다. 이 기준을 만족하지 않는 합성 후보는 최종 데이터셋에서 제외하였다. 따라서 일부 후보가 제거되는 경우가 있더라도, 이는 모델의 오류라기보다 물리적으로 타당하지 않은 데이터를 걸러내는 과정으로 해석할 수 있다.

생성에서 입력은 물리 파라미터 벡터 θ , 온도

T_{amb} , Equivalent Full Cycle (EFC), 경과 시간 t_{cal} , 정규화 cycle 위치 c_{norm} , 프로토콜 특징 벡터 p 로 구성하며, 출력은 cycle 요약 벡터 $\hat{y}$ 로 정의한다.

$$\hat{y} = f_{\phi}(\theta, T_{amb}, EFC, t_{cal}, c_{norm}, p) \quad (4)$$

식 (4)에서 f_{ϕ} 는 학습된 대리모델이며, $\hat{y}$ 는 SoH 관련 요약량을 포함하는 합성 cycle에 대한 요약을 의미한다. 하지만 해당 Stage의 산출물은 최종 산출물이 아닌 합성 후보 집합이다. 해당 후보 집합들은 실측 데이터의 요약 사이의 차이를 보정하는 Stage D로 넘어간다.

2.5 Stage D: 요약 공간 잔차 보정 설계

Stage D에서는 Stage C에서 생성된 합성 요약과 실측 요약 사이의 차이를 줄이기 위해 요약 공간 잔차 보정을 진행한다. 보정 대상은 SoH 관련 요약량의 전반이며, 실측 분위 수 범위를 벗어나는 과도한 이탈을 방지하기 위해, 분위 수 기반 범위 제한을 함께 적용한다.

보정은 기본 합성값에 대한 잔차 예측값과 공분산 구조를 갖는 노이즈를 더한 후 물리적 범위 제한을 적용하는 방식으로 구성하는데, 이는 식 (5)와 같이 나타낸다.

$$y_{cal} = C(y_{\square} + \hat{r}_{\psi}(x) + \varepsilon), \quad \varepsilon \sim N(0, \lambda^2 \Sigma) \quad (5)$$

여기서 y_{base} 은 Stage C에서 생성된 기본 합성 요약량, $\hat{r}_{\psi}(x)$ 는 입력 특징 x 에 대해 학습된 잔차 함수이다. ε 는 잔차 공분산을 반영한다. 잡음 항, $C(\cdot)$ 는 물리적 범위 제한 및 분위 수 기반 clipping 연산을 의미한다. 본 논문에서는 $\lambda = 0.05$, 실측 분위 수 범위 1%-99%로 제한한다.

또한, 보정 이후로 조건 단위 정렬을 추가로 적용하여, 분포 정합성과 물리 일관성을 동시에 개선하도록 설계했다. 해당 Stage D에서는 물리 기반 생성만으로 남을 수 있는 요약 통계 수준의 편차를 줄여, 제한 DoD 조건에서 합성 데이터의 활용 가능성을 높이는 핵심 단계이다.

2.6 Stage E: 품질 관리 게이트 설계

Stage E는 단순히 대량으로 데이터를 생성하기 보다는 실측 데이터의 분포 적합성과 물리 일관성을 동시에 만족하여, 실제 활용할 수 있는 최종 합성 데이터 생성을 진행한다.

해당 품질 평가는 경험적 기준을 적용하며, Area Under the ROC-AUC(Receiver Operating Characteristic Curve) 기반 구분 가능성, KS(Kolmogorov-Smirnov) 통계량 기반 분포 차이, 특징 간 상관 구조 차이, SoH 단조성 위반률, 재시물레이션 drift를 함께 점검하는 방식으로 진행한다. 특히 상관 구조 평가는 모든 요약량을 사용하지만, 보정 이후의 중복적인 변수들이 상관 구조 평가를 과도하게 불안정하게 만드는 줄이기 위해 운전시간, 방전용량, 최소/최대전압, 평균 및 최대 셀 온도 총 6개의 핵심 feature를 사용한다.

재시물레이션 drift는 위 6개의 핵심 특징 중, 최

소 전압과 평균 셀 온도를 제외한 4개에 대해 계산하되, 게이트의 스케일러는 단순 평균이 아닌 전압과 온도 일관성에 더 큰 가중을 부여한 가중 평균으로 집계한다. 이에 대한 가중치는 각각, 0.1, 0.1, 0.4, 0.4로 설정한다. 해당 기준들은 분포에 대한 적합성을 판단하기 위해, 설정값을 적용하여 통과하는 경우, 데이터 생성을 마친다(\$3.2장).

III. 실험 및 결과

제한 DoD 조건에서 제안한 하이브리드 합성 데이터 생성 프레임워크는 이전에서 말한 것과 같이 두 그룹 A, B를 독립적으로 검증한 뒤, 품질 관리 게이트를 통과한 최종 합성 데이터를 병합한 제한 DoD 데이터셋을 구성하는 방식으로 보여준다. 이는 서로 다른 운전 특성을 갖는 제한 DoD 실험군을 단일 분포로 가정하지 않음과 동시에, 일반화할 수 있는 데이터를 구성하기 위한 목적으로 진행된다.

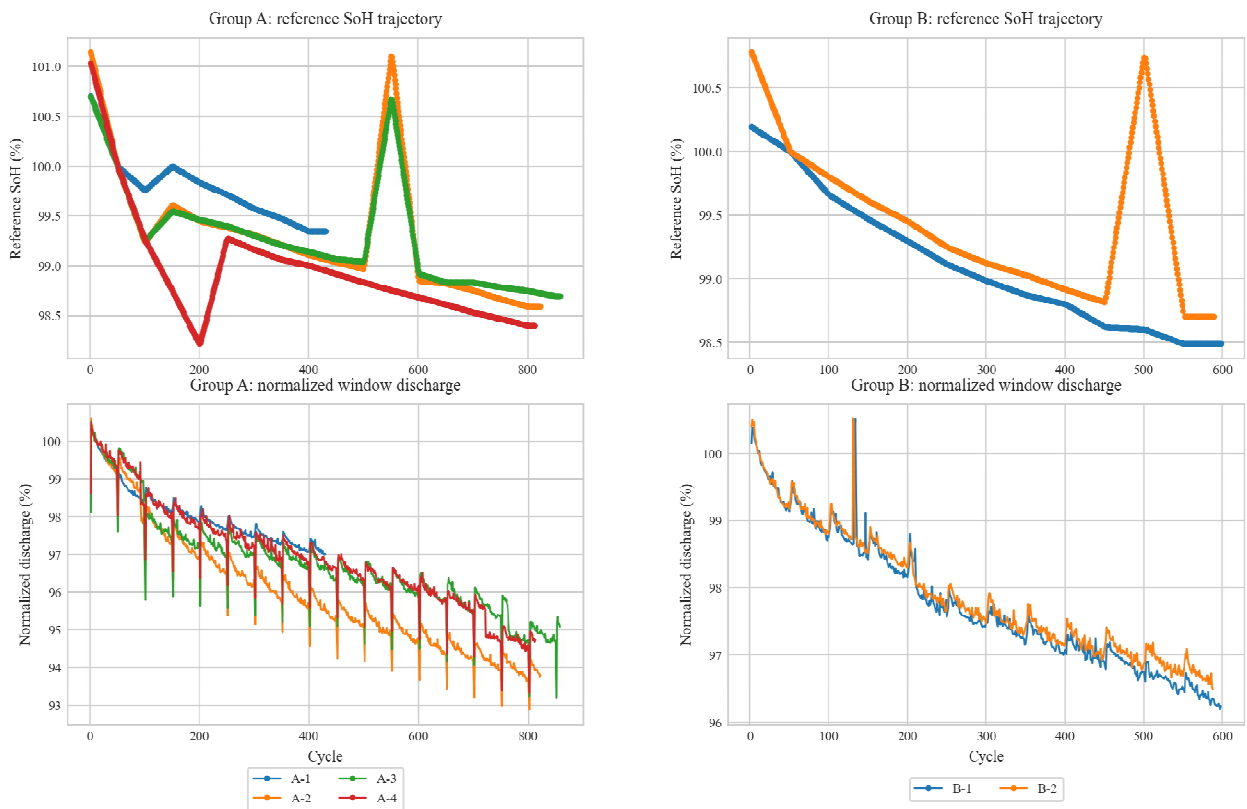


그림 2. 제한 DoD 조건에서 사용된 실측 셀의 참조 SoH 궤적과 정규화된 방전 궤적

Fig. 2. Reference SoH trajectory and normalized discharge trajectory of the measured cell under constrained DoD conditions

3.1 제한 DoD 실측 데이터 특성

실측 데이터는 동일한 제한 DoD 조건에서 획득되었으나, 그룹별 방전용량과 운전시간의 분포는 서로 상이한 양상을 보였다. 그룹 A와 B를 분리한 상태에서 각 Stage 별 진행 결과를 보여준다.

그림 2는 제한 DoD 구간에서 관측된 충·방전 궤적을 바탕으로, Stage A와 Stage B에 활용된 실측 데이터의 참조 SoH 궤적과 정규화된 방전 궤적을 제시한다. 이를 바탕으로 최종 생성된 데이터는 총 375,207개이다.

3.2 Stage E 결과

최종 합성 데이터의 활용 가능성을 검증하기 위해 해당 프레임워크에서는 Stage E의 품질 관리 게이트를 적용했다.

표 2의 결과는 제안한 품질 관리 기준이 제한 DoD 조건에서 실측 데이터와의 차이를 억제하면서도 대규모 합성 데이터셋을 구성하는 데 유효함을 보여준다. 해당 내용에서 각 그룹에서 ROC-AUC는 0.5에 가까울수록 실측 데이터와 합성 데이터의 분포가 구분이 덜 된다는 뜻이다.

표 2. 품질 관리 게이트 결과

Table 2. QC gate results

Evaluation metric	Standard	A	B
ROC-AUC	≤ 0.85	0.7241	0.5782
KS max	≤ 0.30	0.2819	0.0380
Correlation diff.	≤ 0.30	0.2118	0.2552
Monotonic viol.	≤ 0.02	0.0000	0.0000
Drift scalar	≤ 0.20	0.1520	0.1607
Final rows	–	187,929	187,278
QC Pass	T/F	True	True

해당 두 실험군의 품질 지표와 함께 해석하면, 운전 조건 차이가 각 지표에 미치는 영향도 확인할 수 있는데, Stage A에서 확인된 바와 같이 그룹 B는 A보다 방전용량과 운전시간의 중앙값이 더 크고 운전시간 분포도 더 길게 나타났다. 이는 개별 요약

이 안정적으로 정렬되더라도, 용량, 시간, 전압, 온도 SOH 사이의 결합 구조는 더 민감하게 달라질 수 있다.

실제로 그룹 B는 ROC-AUC와 KS max 측면에서는 A보다 우수한 분포 적합성을 보였으나, 평균 상관 차이는 그룹 A보다 다소 크게 나타났다. 이는 B의 합성 데이터가 단변량 분포 수준에서는 실측 자료를 잘 재현했으나, 변수 간 상호의존성까지 완전히 동일하게 복원하는 데에는 상대적으로 더 어려움이 있었음을 보여준다.

3.3 분포 정렬과 합성 데이터 적합성

Stage C에서 곧바로 생성된 합성 후보는 충분한 양을 확보했다. 그러나 Stage C의 출력만으로는 아직 실측 제한 DoD 분포와 정렬되지 않았으며, 특히, 방전용량, 운전시간, 참조 SoH 분포에서 실측과의 차이가 보인다.

이는 Stage D에서 요약 공간 잔차 보정을 도입한 이유를 보여준다. 실제로 보정 이후에는 각 그룹 내에서 실측 분포와 최종 합성 분포의 차이가 유의하게 감소하였다.

Stage D 적용 전후의 실측 대비 오차는 표 3과 같다. 해당 분포에 대해서 B, C와 같은 스테이지를 단순히 거치기만 한다면, pre-D와 같이 큰 오차를 보인다.

표 3. 그룹 A 요약 공간 잔차 보정 미적용/적용 비교

Table 3. Group A: comparison before and after applying summary-space residual calibration

Category	Real	Applying		Error reduction
		pre-D	post-D	
Discharge capacity	1.016	1.867	1.039	97.4%
Duration	4090.0	9450.9	4128.8	99.3%
Reference SoH	99.198	71.132	99.198	100%

그룹 A의 경우, 방전용량 중앙값은 +0.022Ah로, 운전시간 중앙값은 +38.8s의 차이를, 참조 SoH 중앙값 차이는 0.00%로 사실상 일치했다.

표 4와 같이, 그룹 B 또한 방전용량 중앙값이 +0.002Ah, 운전시간 중앙값 -3.6s, 참조 SoH 중앙값 +0.000%로 실측 분포와 가까운 정렬을 진행한다.

해당 그림 3과 같이 병합된 결과와 실측 데이터와의 분포를 확률 밀도(Density)와 누적 분포 함수(ECDF) 기준으로 분포 형상을 비교한다.

표 4. 그룹 B 요약 공간 잔차 보정 미적용/적용 비교
Table 4. Group B: comparison before and after applying summary-space residual calibration

Category	Real	Applying		Error reduction
		pre-D	post-D	
Discharge capacity	1.491	1.721	1.493	99.3%
Duration	8366.3	9059.2	8362.7	99.5%
Reference SoH	99.171	60.310	99.171	100%

3.4 순수 물리 기반 시뮬레이션 비교

표 5에서는 제안 프레임워크 (Ours)를 순수 물리 기반 생성 (Physics-only)과 일반 데이터 기반 생성

기인 Generative Adversarial Networks (GAN), Variational Autoencoder (VAE)와의 비교를 진행한다.

제안된 프레임워크는 두 그룹 모두에서 KS max와 Correlation diff.가 가장 낮거나 가장 안정적으로 나타났다. ROC-AUC의 경우, GAN이 제일 좋게 나왔으나, 종합 해석 결과로는 단변량 분포 형상인 KS max나, 변수 간 구조 보존 측면인 Correlation diff.에서 극단적으로 높은 값을 보이며 GAN이 분류기를 속이는 방향으로 해석할 수 있다. 따라서, 제한 DoD 합성 데이터의 품질은 단일 지표가 아닌 복수 지표를 종합하여 해석할 필요가 있다.

표 5. 제한 DoD 조건 생성 방식별 합성 데이터 품질 비교
Table 5. Synthetic Data Quality Comparison Across Generative Methods in Restricted-DoD Conditions

Metric	Group A			Group B		
	Ours	GAN	VAE	Ours	GAN	VAE
ROC-AUC	0.7241	0.6613	0.5818	0.5782	0.5548	0.7392
KS max	0.2819	0.5416	0.5000	0.0380	0.9854	0.5000
Correlation diff.	0.2118	0.4720	1.0000	0.2552	0.3615	1.0000

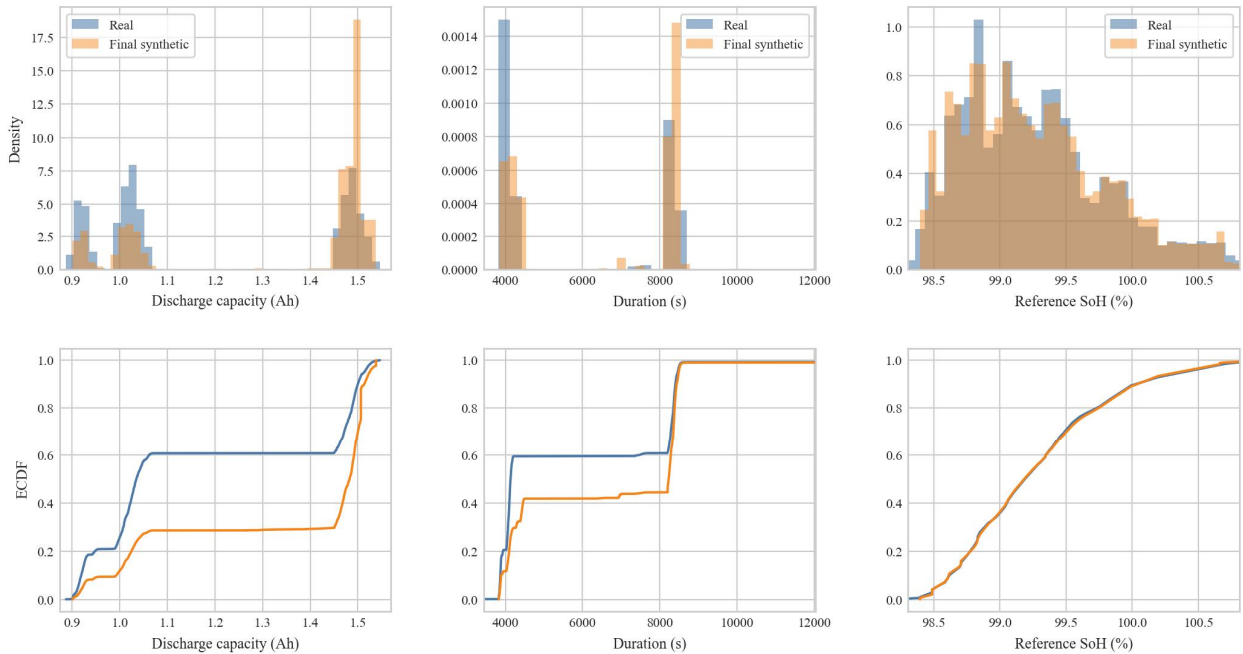


그림 3. 실측 제한 DoD 데이터와 최종 합성 데이터의 분포 비교. 병합 데이터셋 기준 Empirical Cumulative Distribution Function (ECDF) 제시 및 정량 해석은 각 그룹별 요약

Fig. 3. Distribution comparison between measured constrained DoD data and the final synthetic data. ECDF of the merged dataset is presented, with quantitative summaries provided for each group

3.5 방법론 성능

외부 공개 데이터셋에서 본 논문의 목적과 완벽하게 정합하는 데이터는 존재하지 않는다. 외부 적용 가능성을 검토하기 위해, 전주기 충·방전 데이터셋인 NASA FY08Q4와 제한 DoD 조건에 근접한 Stanford SECL의 W8 및 W10 시나리오를 대상으로 제안 방법을 적용하였다. 이때 원 데이터와 재구성 결과 간의 해석 차이를 최소화할 수 있는지를 중점적으로 평가하였다.

표 6에서는 본 논문의 설계와 동일하게 생성하였을 때, Median relative error max과 KS max 두 지표 모두 각각 0.000087, 0.0141이 가장 큰 값일 정도로 매우 작았다. 이는 제안 방법이 내부 데이터셋에 과도하게 특화되지 않고, 외부 데이터셋에 대해서도 일정 수준의 재구성 안정성을 보일 수 있음을 시사한다.

표 6. 외부 데이터셋의 방법론 적용 성능 요약

Table 6. Performance summary of the proposed method on external datasets

Dataset	Scope coverage	KS max	Median relative error max
NASA FY08Q4	0.588	0.010554	0.000087
Stanford SECL W8	0.602	0.014108	0.000005
Stanford SECL W10	0.608	0.013684	0.000002

IV. 결론 및 향후 과제

본 연구는 제한 DoD 조건에서 부족한 실측 열화 데이터를 보완하기 위한 하이브리드 합성 데이터 생성 프레임워크를 제안했다. 6개 셀, 4,027개 cycle 요약으로부터 375,207개의 합성 데이터를 생성했고, 분포 정합성을 확인했다. 또한 외부 공개 데이터셋에 대한 적용 결과, 제안한 하이브리드 프레임워크는 제한 DoD 조건과 유사한 환경에서도 낮은 분포 차이를 유지하여 일정 수준의 외부 적용 가능성을 보였다.

해당 데이터에서는 실측 대비 방전용량, 운전시간, 참조 SoH의 중앙값 차이는 그룹 A에서

+0.022Ah, +38.8s, 0.000%가, 그룹 B에서는 +0.002Ah, -3.6s, 0.000%로 나타났다. 사전정의된 품질 게이트 또한 통과하여, 실측 데이터와 큰 차이를 줄였다.

향후에는 다양한 DoD 조건에 정합한 데이터셋을 추가로 확보하고, 실측 데이터 기반의 물리 파라미터 복원 정확도를 높이기 위한 장주기 실험 데이터를 확충할 필요가 있다. 본 논문은 이를 통해, 제한 DoD 조건에서의 실제 운용되는 배터리 데이터 부족 문제를 다루는 일반적인 합성 데이터 구축 방법론으로 발전할 수 있다.

References

- [1] J. Vetter, P. Novák, M. R. Wagner, C. Veit, K.-C. Möller, J. O. Besenhard, M. Winter, M. Wohlfahrt-Mehrens, C. Vogler, and A. Hammouche, "Ageing mechanisms in lithium-ion batteries", *Journal of Power Sources*, Vol. 147, No. 1-2, pp. 269-281, Sep. 2005. <https://doi.org/10.1016/j.jpowsour.2005.01.006>.
- [2] A. Barré, B. Deguilhem, S. Grolleau, M. Gérard, F. Suard, and D. Riu, "A review on lithium-ion battery ageing mechanisms and estimations for automotive applications", *Journal of Power Sources*, Vol. 241, pp. 680-689, Nov. 2013. <https://doi.org/10.1016/j.jpowsour.2013.05.040>.
- [3] H. de Vries, T. T. Nguyen, and B. Op het Veld, "Increasing the cycle life of lithium ion cells by partial state of charge cycling", *Microelectronics Reliability*, Vol. 55, No. 11, pp. 2247-2253, Nov. 2015. <https://doi.org/10.1016/j.microrel.2015.08.014>.
- [4] S. Jenu, A. Hentunen, J. Haavisto, and M. Pihlatie, "State of health estimation of cycle aged large format lithium-ion cells based on partial charging", *Journal of Energy Storage*, Vol. 46, Art. no. 103855, Feb. 2022. <https://doi.org/10.1016/j.est.2021.103855>.
- [5] M. Lin, Y. You, J. Meng, W. Wang, J. Wu, and D.-I. Stroe, "Lithium-ion battery degradation

- trajectory early prediction with synthetic dataset and deep learning", *Journal of Energy Chemistry*, Vol. 85, pp. 534-546, Oct. 2023. <https://doi.org/10.1016/j.jechem.2023.06.036>.
- [6] H. Jin, R. Ru, L. Cai, J. Meng, B. Wang, J. Peng, and S. Yang, "A synthetic data generation method and evolutionary transformer model for degradation trajectory prediction in lithium-ion batteries", *Applied Energy*, Vol. 377, Part D, pp. 124629, Jan. 2025. <https://doi.org/10.1016/j.apenergy.2024.124629>.
- [7] H. Zhang, F. Altaf, and T. Wik, "Synthetic dataset of LG M50 batteries with different degradation pathways", *Data in Brief*, Vol. 57, pp. 111076, Dec. 2024. <https://doi.org/10.1016/j.dib.2024.111076>.
- [8] W. Li, H. Samsukha, B. van Vlijmen, L. Yan, S. Greenbank, S. Onori, and V. Viswanathan, "Fast data augmentation for battery degradation prediction", *Energy and AI*, Vol. 21, Art. no. 100542, Sep. 2025. <https://doi.org/10.1016/j.egyai.2025.100542>.
- [9] M. Hassanaly, P. J. Weddle, R. N. King, S. De, A. Doostan, C. R. Randall, E. J. Dufek, A. M. Colclasure, and K. Smith, "PINN surrogate of Li-ion battery models for parameter inference, Part I: Implementation and multi-fidelity hierarchies for the single-particle model", *Journal of Energy Storage*, Vol. 98, Part B, Art. no. 113103, Sep. 2024. <https://doi.org/10.1016/j.est.2024.113103>.
- [10] M. Hassanaly, P. J. Weddle, R. N. King, S. De, A. Doostan, C. R. Randall, E. J. Dufek, A. M. Colclasure, and K. Smith, "PINN surrogate of Li-ion battery models for parameter inference, Part II: Regularization and application of the pseudo-2D model", *Journal of Energy Storage*, Vol. 98, Part B, Art. no. 113104, Sep. 2024. <https://doi.org/10.1016/j.est.2024.113104>.
- [11] A. Thelen, Y. H. Lui, S. Shen, S. Laflamme, S. Hu, H. Ye, and C. Hu, "Integrating physics-based modeling and machine learning for degradation diagnostics of lithium-ion batteries", *Energy Storage Materials*, Vol. 50, pp. 668-695, Sep. 2022. <https://doi.org/10.1016/j.ensm.2022.05.047>.
- [12] R. G. Nascimento, M. Corbetta, C. S. Kulkarni, and F. A. C. Viana, "Hybrid physics-informed neural networks for lithium-ion battery modeling and prognosis", *Journal of Power Sources*, Vol. 513, Art. no. 230526, Nov. 2021. <https://doi.org/10.1016/j.jpowsour.2021.230526>.
- [13] V. Sulzer, S. G. Marquis, R. Timms, M. Robinson, and S. J. Chapman, "Python Battery Mathematical Modelling (PyBaMM)", *Journal of Open Research Software*, Vol. 9, No. 1, Art. no. 14, Jun. 2021. <https://doi.org/10.5334/jors.309>.
- [14] B. Planden, N. E. Courtier, M. Robinson, A. Khetarpal, F. Brosa Planella, and D. A. Howey, "PyBOP: A Python package for battery model optimisation and parameterisation", *Journal of Open Source Software*, Vol. 10, No. 116, pp. 7874, Dec. 2025. <https://doi.org/10.21105/joss.07874>.
- [15] S. G. Marquis, V. Sulzer, R. Timms, C. P. Please, and S. J. Chapman, "An Asymptotic Derivation of a Single Particle Model with Electrolyte", *Journal of The Electrochemical Society*, Vol. 166, No. 15, pp. A3693-A3706, Nov. 2019. <https://doi.org/10.1149/2.0341915jes>.
- [16] C.-H. Chen, F. Brosa Planella, K. O'Regan, D. Gastol, W. D. Widanage, and E. Kendrick, "Development of Experimental Techniques for Parameterization of Multi-scale Lithium-ion Battery Models", *Journal of The Electrochemical Society*, Vol. 167, No. 8, Art. no. 080534, Jan. 2020. <https://doi.org/10.1149/1945-7111/ab9050>.

저자소개

최 승 민 (Seungmin Choi)



2026년 2월 : 국립금오공과대학교
컴퓨터공학과(공학박사)
2026년 3월 ~ 현재 :
국립금오공과대학교
컴퓨터·AI융합공학과 석사과정
관심분야 : 자연어 처리, 지식그래프,
거대언어모델, AI응용

김 광 섭 (KwangSeop Kim)



2020년 2월 : 포항공과대학교
전자전기공학과(공학박사)
2019년 12월 ~ 현재 :
국방과학연구소 선임연구원
관심분야 : 군용 배터리,
전력변환장치

최 현 준 (HyunJun Choi)



2020년 8월 : 울산과학기술원
(UNIST) 전기및전자공학과
(공학박사)
2020년 7월 ~ 2022년 8월 :
LG전자 선임연구원
2022년 9월 ~ 2024년 8월 :
한국전자기술연구원(KETI)

선임연구원

2024년 9월 ~ 현재 : 국립금오공과대학교 전자공학부
반도체시스템전공 조교수

관심분야 : Power Electronics, Fault Diagnosis

박 범 용 (Bumyong Park)



2015년 8월 : 포항공과대학교
전자전기공학과(공학박사)
2015년 9월 ~ 2017년 8월 :
삼성전자 생산기술연구소
책임연구원
2017년 9월 ~ 2022년 8월 :
금오공과대학교 전자공학부

조교수

2022년 9월 ~ 현재 : 금오공과대학교 전자공학부 부교수

관심분야 : Robust control, robot manipulator system,
신호 처리, 임베디드 제어 시스템

정 유 철 (Yuchul Jung)



2011년 2월 : 한국과학기술원
(KAIST) 전산학과(공학박사)
2009년 1월 ~ 2013년 7월 :
한국전자통신연구원 (ETRI)
선임연구원
2013년 8월 ~ 2017년 8월 :
한국과학기술정보연구원(KISTI)

선임연구원

2017년 8월 ~ 2022년 2월 : 국립금오공과대학교
컴퓨터공학과 조교수

2022년 2월 ~ 현재 : 국립금오공과대학교 인공지능공학과
부교수

관심분야 : 거대언어모델, 자연어처리, 지식그래프,
한국어 음성 인식/합성, AI응용, 배터리 SOH예측